An Empirical Study On The Nature Of Corruption Amongst Nigerian Firms: Causes, Channels And Detection

Omofolamihan Olaboye Malomo

Thesis submitted to The University of Sussex
for the degree of Doctor of Philosophy

May 2014

Declaration

I hereby declare that this thesis has not been previously submitted in whole or in part to this or any other University for the award of any other degree.

Signature:.........................

Omofolamihan Malomo

UNIVERSITY OF SUSSEX

Omofolamihan Olaboye Malomo, Doctor of Philosophy

An Empirical Study On Corruption Amongst Nigerian Firms: Causes, Channels And Detection

Summary

This dissertation discusses the economic issues surrounding corruption at the firm-level in Nigeria with a specific focus on bribery. This involves an analysis of the paying and reporting of bribes by firm managers in Nigeria. The first chapter uses data from two business surveys to explain the determinants of the incidence of bribery and the magnitude of bribes, respectively. A two-stage analysis is conducted to test for the independence of the processes determining the incidence and the magnitude of bribery. The results show that the propensity to bribe is determined by required meetings with public officials while the size of bribe is driven by firm profitability indicators. The second chapter tests the reliability of methods used to ask individuals sensitive questions on different forms of business malpractice. Indirect methods are tested against the randomised response method. The indirect method protects the managers from stigmatisation by asking them about the behaviour of an agent representative of themselves; the randomised response method asks the interviewee to base their response on the result of a private coin-toss. The results show that the indirect method produces higher and more plausible estimates of wrongdoing than the randomised response method. The third chapter investigates why the randomised response method sometimes fails in eliciting honest responses from sensitive questions despite assuring the managers of anonymity. The roles of trust in the interviewer and the probability of detection are considered along with other potential explanations. Results indicate that lack of trust and the fear of detection are associated with underreporting of sensitive acts. The final chapter examines the relationship between bribery and ethnic networks. The ethnicities of the managers and their local political representatives are used to measure ethnic networks. Results show that co-ethnic firm managers are less likely to pay a bribe than non co-ethnics. Also, there is a positive association between ethno-linguistic fractionalisation and bribery which, in fractionalised areas, eradicates the negative effect of co-ethnicity on bribery.

Acknowledgements

I am grateful to God for my life and for giving me the strength to complete this dissertation.

I would like to thank my supervisors Professor Alan Winters and Professor Richard Dickens for their advice and support. I would also like to thank Professor Barry Reilly and Mike Barrow for comments on my work. I would also like to express my gratitude to the doctoral researchers at Sussex, including Edgar Cooke, Farai Jena, Janani Akhil, Grazia Pacillo; whose friendship made the PhD process much easier. I am also grateful to Dr. Michael Sulu for his encouragement and advice.

I would also like to express my thanks to employees of the EFCC; ICPC; and Lambeth Palace Library for their helpful assistance in my research. My thanks also go to Mr. Folu Olamiti; Ms. Lilian Ekeanyanwu; Mr. Emmanuel Ojo; Ms. Olamide Adubiagbe; Mr. Olayinka Babatunde; and Mr. Sina Babasola for their assistance in my research.

I would also like to acknowledge the ESRC for the financial support which allowed me to complete my research. I would further like to acknowledge the help and support given to me by the staff within the school of Business, Management and Economics.

I would like to thank the participants of the Economics of Corruption Course; EUDN PhD Workshop; Sussex DPhil Lunch Seminar; Sussex Research In Progress Seminar; Royal Economic Society PhD Presentation Meeting; Sussex Annual DPhil Conference; PACDEV conference; CSAE Conference; SES Conference; SMYE Conference; ESRA Conference and EEA|ESEM conference for their comments on my research

Finally, I would like to thank my family for their confidence and continuous support throughout this process.

Contents

List of Tables

List of Figures

1 Introduction

This dissertation discusses the economic issues surrounding corruption at the firm-level in Nigeria with a specific focus on bribery. Much of the conceptual and empirical research on corruption looks at it from the point of view of the corrupt transaction, where a government official acts based on a cost-benefit analysis of the gains from accepting a bribe and the expected punishment of being caught [Becker , 1968]. While the current dissertation allows for this mechanism to take place, it also considers the influence of market forces on the level of corruption. With the latter view, potentially corrupt government officials behave like bribe maximising agents, and the incidence of bribery is determined by the structure of the market, the demand for the services of the government official, and the degree to which officials can collude in collecting bribes [Shleifer & Vishny , 1993, Olken & Barron , 2009].

The current study investigates the influence of market forces on the prevalence of bribery. This is studied in the context of bribes paid by firm managers to government officials and local politicians in Nigeria. Firm managers make a variety of informal payments, including payments to the police; customs officials; and court officials, to speed up the process of regulation, to access information, to win a court case, and to avoid fines for illegal behaviour. To investigate these payments, this study uses a conceptual framework, that was introduced by Svensson [2002], to describe the forces at work when a firm bribes an official; and the determinants of the amount of bribe paid in such transactions. Data is taken from 2 waves of the World Bank Enterprise Survey for Nigeria, conducted in 2007/8 and 2008/9. In addition to this, data from the Nigeria Bureau Of Statistics and the Economic and Financial Crimes Commission is used to complement the World Bank Dataset. Also, primary data, concerning characteristics of firm managers, was collected from a sample frame of 2,110 establishments in Nigeria.

Corruption is defined as the use of public office, or power, for private gain [Leite & Weidemann , 2002]; bribery is the paying of informal gifts or money to speed up the regulatory process or to bypass regulation [Laws Of The Federation Of Nigeria , 1990]. This dissertation makes many contributions to the body of knowledge; some of the major themes that are present throughout this work include: the measurement of bribery (Chapters 3, 4 and 5); the data on bribery (Ibid.); the explanation for the underreporting of bribery (Chapter 5); and the channels through which bribery operates (Chapters 3 and 6). The four main chapters of this work discuss: the factors that determine which firms will pay a bribe and how much they will pay; the preferred ways of measuring bribery (specifically) and corruption (in general) at the firm level; why firms lie when asked sensitive questions in surveys when given the assurance of anonymity by the interviewers; the effect of ethnic diversity on the the proportion of bribing firms in a region; and the effect of a (mis)match between the ethnicity of firm managers and the ethnicity of local government official on the propensity for a bribe to be paid by the firm.

This work fills many gaps in the literature. Despite the growing interest in the economics of corruption; relatively little has been done on corruption at the firm level compared to corruption at the country level. Part of the reason for this is the relative lack of data on corruption amongst firms; this comes from the potential to underreport due to the sensitive nature of the data [Reinikka & Svensson , 2003]. This study fills this gap by focusing on firms in Nigeria while using information from the wider economy and the issues involving corruption within Nigeria.

Another gap that is filled by this work is the relative lack of research performed on corruption in Africa, and specifically Western Africa. One explanation for this is the lack of data for the entire region compared with Western Europe and Northern America. An outcome of this is that very little work has been done on firm-level corruption within a Western African context. It is important to understand whether corruption in this region follows the same processes as corruption elsewhere in the world.

Also, there has been a need for a disaggregation of corruption indices to focus on the sub-national regional variation in the prevalence of corruption; and the inter-industry variation in the prevalence of corruption. Whilst a country might be perceived to be more corrupt than another, it is possible that some regions and industries within the more corrupt country are less corrupt than some regions and industries in the less corrupt economy. By focusing solely on country-level data, some potentially useful information is lost. This study seeks to retrieve this information and add it to the body of knowledge. This dissertation uses both parametric and non-parametric methods to analyse corruption at the firm level.

A running theme throughout this work is that useful information can be retrieved from firms when asking them about their experiences of bribery. Furthermore, information can be revealed by firms by observing their reporting behaviour when using different techniques to ask sensitive questions [Iarossi , 2006]. When asking firms about their behaviour, some methods are preferred to others. For example, this study finds that asking firm managers if they have ever paid a bribe indirectly provides more useful information compared with asking them via the randomised response technique. Essentially the strength of the false-consensus effect helps to overcome social desirability bias in the case of indirect questions.

The thesis also comments on the different channels through which corruption can occur. These include: meetings with government officials [Svensson , 2003]; and ethnic cleavages [Mauro , 1995]. A running sub-theme in the analysis is that willingness to pay and ability to pay have seperate effects on the payment of bribes and the amount of bribe paid, respectively. Willingness to pay seems to affect the payment of bribes, whilst ability to pay seems to influence the amount of bribe paid, conditional on a bribe being paid.

The analysis is applied to Nigeria for many reasons. In order to carry out this study, it was necessary to use an economy where corruption exists. Nigeria has consistently ranked in the bottom quartile of Transparency International's Corruption Perceptions Index and has had a number of high-profile corruption cases. This indicates that corruption is present within the Nigerian economy.

Secondly, Nigeria has Africa's largest population, with over 160 million inhabitants. Coupled with a 0.2 share of Sub-Saharan Africa's 2011 GDP, it represents a relatively large nation with an economy that has a non-trivial impact on African and World GDP. In the year 2014 its GDP was calculated to be $510 billion; making it Africa's largest economy. Recent policies have aimed to reduce the overall level of corruption within the country by: introducing due process guidelines; signing up to the Extractive Industries Transparency Initiative (EITI); and promoting the Economic And Financial Crimes Commission (EFCC) and Independent Corruption Practices Commission (ICPC).

Results from previous research suggest that corruption is inevitable whenever a country transitions from one type of political regime to another [Huntingdon , 1968]. For example, a country moving from an autocracy to democratic rule might see many people vying for power and rents. Nigeria makes for a potentially interesting study since it recently transitioned from military rule to democracy in 1999. It represents a case study of the nature of corruption amongst firms in a relatively young democracy. The current study also pays attention to the formal manufacturing industry in Nigeria. This industry contributed to 4% of Nigerian Gross Domestic Product; and 8% of GDP growth in 2012.

Summary

The first main chapter of this dissertation investigates the factors that determine whether or not a firm will pay a bribe; and how much bribing firms pay. This chapter applies a simplified version of the Svensson [2003] model to data on Nigeria using a larger sample; a larger vector of explanatory variables; and more disaggregated variables. These three additions allow for better conclusions to be made about the supply of bribes in Nigeria. The data for this chapter comes from two sources: the World Bank Enterprise Survey (ES); and the Nigeria Bureau Of Statistics & Economic And Financial Crimes Commission Business Survey On Crime And Corruption (NBS). These surveys use different methods to ask firms about bribery. The ES survey asks the bribery-related questions indirectly whilst the NBS survey asks the questions directly.

The conceptual framework; and diagnostic tests on the data suggest the usefulness of the two-part model in analysing the data. Nevertheless, findings from a Heckman-selection model and a censored Tobit model are also presented. The results indicate two separate processes driving the propensity to pay a bribe; and the amount of bribe paid, respectively. Encounters with government employees are associated with an increased propensity to pay a bribe whilst current and expected future profitability influence the amount of bribe that is paid. Firms that are more willing to pay a bribe, i.e. those who are required to deal with one or more government officials, are more likely to pay a bribe than firms who are not. Amongst bribing firms, those who are more able to pay, i.e. those with a higher current and expected future profit, generally tend to pay larger bribes. These results remain when using different datasets; different econometric specifications; and different measures of firm level bribery.

The second chapter of this dissertation investigates the preferred methods of asking firm managers about bribery and illegal behaviour. Three methods are identified and two methods are tested against each other. The three main methods are: direct questioning (DQ); indirect questioning (IQ); and randomised response questioning (RRQ). This chapter fills a gap in the literature by comparing the two methods which had previously not been compared against each other: IQ & RRQ.

The IQ method of asking sensitive question uses a projective technique by asking about the behaviour of similar firms. This technique borrows from the results of the false consensus effect, which finds that people who engage in sensitive behaviour are more likely to overestimate the prevalence of that behaviour amongst their peers. The inference from this is that firm managers who bribe are more likely to report that similar firms also bribe.

The RRQ method of asking sensitive questions attaches the response of the firm manager to a probability distribution. This is done in private using a randomising device: a fair coin. Interviewees are asked to base their response to the sensitive question on the result of a coin toss. Interviewees are told to reveal their true status only if a specified random event (a heads) occurs, the result of which is unknown to the interviewer. This procedure helps to remove any social desirability bias from the sensitive question because the interviewer cannot connect a “yes” response to guilt on the part of the interviewee.

Parametric and non-parametric methods are used to compare the IQ and RRQ methods. The respective questioning strategies are tested based on the rate of admittance of guilt that they are each able to generate. In order to exploit the nature of the RR design the responses to RRQs are modelled as binary responses to a sensitive question with the possibility of false-negatives. Results show that the IQ seems to perform better in getting honest responses from sensitive questions. The RRQ seems better suited to estimating the level of misreporting behaviour, labelled “reticence”, amongst firm managers.

The third chapter seeks to explain why firm managers display reticence in RRQs despite the statistical assurance of anonymity, and the impossibility of detection.

The role of trust is brought up. It is argued that a lack of trust amongst firm managers in general might indicate a propensity for the firm manager to distrust an interviewer. If this is the case then the manager might not trust the interviewer to keep the results anonymous. Alternatively, the manager might believe that the technique is a ploy to trick them and might not trust the RRQ process.

Another reason for the observed reticent responses might be the fear of detection. For example, regardless of the assurances of anonymity, the industry and region that the firm is located in was not expected to be kept confidential. This allows for a positive, albeit small, probability of detection, which would decrease with the size of the industry. So a monopoly, or a firm with very few competitors in its industry-location cell, might fear detection and lie about bribing much more than a firm with a heavily populated cell.

This study uses the following variables as indicators of latent trust: the percentage of material orders paid for after delivery (instead of: before delivery; or on delivery); the percentage of sales orders paid for before delivery (instead of: on delivery; or after delivery); and the percentage of orders that were written (instead of: oral with witness; or oral without witness). These are used as indicators of trust within the business environment and might relate to general trust (including trust in the RRQ/interview process). This chapter measures the fear/probability of detection using the number of other firms in the industry-location cell; and the self-reported number of close competitors.

Results show that trust and a higher number of self reported competitors are associated with a lower probability of reticence, giving credibility to the story about trust and the fear of detection.

The fourth chapter seeks to address another factor influencing bribery: ethnic ties. Some work has been done on the relationship between ethnic fractionalisation and corruption; however, relatively little has been researched on ethnic diversity and bribery at the micro (firm) level. This is important because there are a number of different situations that can allow for bribery to show up.

Firm managers who are of an ethnic minority might be discriminated against and asked to pay larger bribes; and more frequent bribe payments, in this case firms in ethnically homogeneous areas will pay lower bribes; and bribe less often. Alternatively firm managers of the majority ethnic group might be more willing to pay bribes to their own kin group, in this case both the frequency and size of bribe payments increase with ethnic homogeneity. These two phenomena can also show up at the individual level, regardless of the level of diversity that occurs at the regional level. Also, this chapter investigates the use of language and bi-lingualism in avoiding bribe payments.

The chapter makes multiple contributions. Firstly, it constructs four different indices of corruption, at the national and sub-national level, to describe the different measures of corruption that have been conceptualised by the literature (See Appendix A.4.3). It then constructs measures of linguistic, religious and ethnic fractionalisation & polarisation for Nigeria; something which has not been done recently due to the unavailability of data. The diversity indices are constructed using primary and secondary data. The measures of linguistic diversity take into account the possibility of bi-lingualism, something that has not been done before. Bi-lingualism is important to account for because a minority, if discriminated against, might be able to avoid or reduce the impact of this discrimination by speaking the same language as the majority group, thereby gaining rapport and avoiding having to make a bribe payment. A study which attached a person's language to their ethnic group would ignore the possibility of multiple languages being spoken. This chapter overcomes this restriction and allows for a plethora of language indices to be constructed using the methods proposed in the literature [Greenberg , 1956]. A regional level analysis is conducted and results show no significant effect of diversity on any of the corruption indices at the state level; however, significant result show up when looking at the local government level.

In summary, this dissertation uses both primary and secondary data. It constructs indices of corruption and ethnic; linguistic; and religious diversity for Nigeria at the National; state; and local government levels, respectively. It analyses the regional and industry sub-sector variation in the prevalence and magnitude of bribery amongst formal manufacturing firms in Nigeria. It provides arguments in favour of indirect questioning over randomised response questioning when asking sensitive questions; and also provides explanations for the misreporting of data when using randomised response questioning. The thesis includes the idea that useful data can be retrieved in the field of corruption by using specific techniques and that this data can be used to conduct economic analysis at the firm-level in developing countries.

The rest of this dissertation is organised as follows. The next chapter (Chapter 2) describes the data that is used in the subsequent chapters. Chapter 3 presents the analysis of the factors influencing: the propensity to pay a bribe; and the amount of bribe paid, respectively. This chapter uses direct and indirect questions on bribery to ask firms about their behaviour. The validity of indirect questions is investigated in Chapter 4. This contribution compares the usefulness of indirect questioning in comparison with randomised response questioning, concluding that the former is preferred to the latter in its ability to achieve a higher rate of truth-telling on the part of interviewees. Chapter 5 asks why the method of randomised response questioning often fails to achieve its intended purpose of getting honest answers to sensitive survey questions. Different explanations are presented and tested. Results show that trust and detection seem to be useful explanations for the observed state of affairs. Chapter 6 uses data from the survey asking bribe-related questions directly; and the analysis of Chapter 3 to analyse the link between ethnic networks and corruption. Finally Chapter 7 discusses the contributions of this dissertation; concludes; and proposes extensions for future research. Thus, Chapter 3 applies a methodology that has been used in the literature to a new setting: Nigeria. Chapter 4 tests this methodology against an alternative; in doing this it raises the question of why, despite being guaranteed anonymity, firm managers are reticent in their responses to sensitive questions. Chapter 5 seeks to answer this question by investigating the roles of trust; detection; and other firm characteristics. Chapter 6 uses new data to extend the work of Chapter 3 by looking at the role of ethnicity in business outcomes. In summary, Chapter 3 presents the framework that is used in describing bribery amongst firms; Chapters 4 and 5 justify the use of the methodology that is used in Chapter 3. Given this justification, Chapter 6 proceeds to extend the framework of Chapter 3 by including more information to the analysis. Chapter 7 discusses the results of the main chapters and evaluates the contribution of this dissertation to the body of knowledge.

1.1 Corruption At The Firm Level In Nigeria: Comparative Data

An international comparison of indirect costs amongst firms in Brazil, China, India, Indonesia, Kenya, Nigeria and South Africa showed that, in all countries, bribes account for 2% of a firms total annual sales (Figure 1). Despite this, while 25% of firms in Nigeria perceive corruption to be a serious problem, a higher percentage of firms in other countries deem corruption to be a serious obstacle to business operations [Iarossi, Mousley & Radwan , 2009] [1].

Other sources which seek to quantify the impact of corruption on the business environment also seem to confirm the idea that although corruption is perceived as a problem in Nigeria, it is not significantly worse than other countries. Transparency International's Corruption perception index ranks the degree of corruption in Nigeria as being similar to that of Indonesia, whilst the corruption levels in Kenya and Venezuela appear to be worse (Table 1.1).

A potential explanation for these results is that firm managers become accustomed to the corruption in the business environment and therefore report lower levels of it despite the problem being larger than perceived. To rule out this possibility the current study uses objective measures of corruption such as the amount

[1] 11 out of Nigeria's 37 geo-political states were used for this comparison.

Figure 1: Indirect Costs Of Manufacturing Firms For Different Countries In The Year 2007

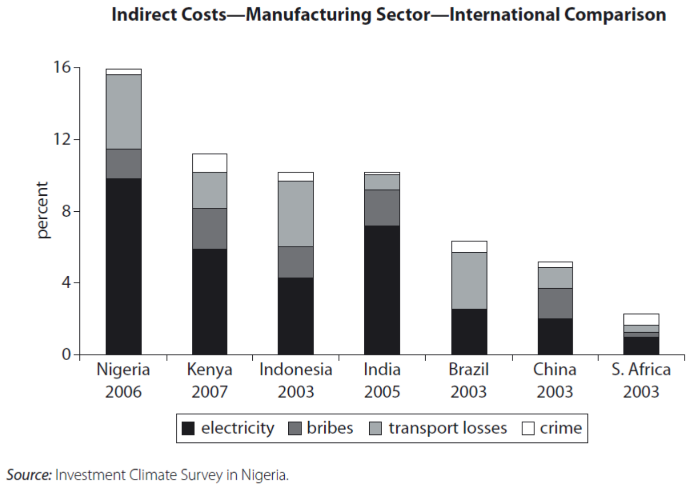


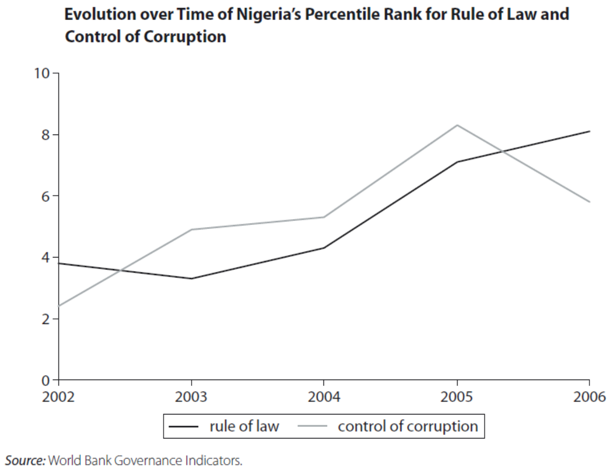
Table 1.1: 2007 Corruption Perceptions Index

Country	Rank (out of 180 countries)	Index
South Africa	43	5.1
Brazil	72	3.5
China	72	3.5
India	72	3.5
Indonesia	143	2.3
Nigeria	**147**	**2.2**
Kenya	150	2.1
Venezuela	162	2.0

The corruption perceptions index ranges between 0 (highly corrupt) and 10 (highly clean). In the 2007 CPI Nigeria ranks 147th out of 180 countries.

of informal payments made to conduct business operations. When using these objective measures the amount of bribes paid in Nigeria is on average lower than the amount paid in Kenya and similar to that of Indonesia [Iarossi, Mousley & Radwan , 2009]. Looking at the changes in the levels of corruption over time shows a sustained decrease in the level of corruption over the past decade (Figure 2). Some of these changes have been attributed to the enacting of the Corrupt Practices Act and the establishment of anti-corruption agencies such as the Economic And Financial Crimes Commission (EFCC) and Independent Corrupt Practices and Other Related Offences Commission (ICPC). Thus, while corruption appears to be a problem within Nigeria, respondents seem to consider the efforts to reduce corruption to be effective and expect the problem to become less of an obstacle to business in the future [Transparency International , 2007], evidence for this is provided in the improving CPI and governance indicators.

Figure 2: Evolution Of Governance And Corruption Indices For Nigeria Over Time



2 Data On Firm Level Bribery In Nigeria

2.1 Introduction

This investigation uses primary and secondary data. The primary data is collected, by the author, by directly contacting 2,110 firm managers via a telephone interview. The secondary data is collected from research, government, and non-governmental organisations including, but not restricted to: the Nigeria Bureau of Statistics; the World Bank; the Economic And Financial Crimes Commission; and the Centre For Law Enforcement Education. Many online sources were also used, including: City Population, Ethnologue, and the World Bank Development Indicators. Secondary data that is used in the analysis include: a survey of the names of local government chiefs; collection of spatial data on the local government areas; and other data. This data was collected in order to get an insight into the business environment faced by firms in Nigeria. Some of the uses of the data include: assigning firms to their respective local government areas; creating estimates of ethnolinguistic fractionalisation for the different regions of the country; and estimating the population of different ethnic groups within the country.

The fourth main chapter of this dissertation (Chapter 6) uses information on the ethnic background of managers to investigate a possible link between ethnic networks and bribe payments between firm managers and local politicians. Fieldwork was carried out in order to get this information. The names of the company owners was available from the Corporate Affairs Commission, the body that regulates the formation and management of companies. The initial data collection strategy was to convert the names of the firm managers into their respective ethnic groups. However, this information was not available in a format accessible to the public; and extracting this information from would require a substantive number of worker-hours. Some of the larger firms, and firms with more than one branch in Nigeria, have information about their managers on the internet, however, the number of these firms was very small. In order to get the required information the author contacted the respective 2,110 firms individually by telephone. Firms were asked for information about the manager in 2007. Data on 1,267 firms was usable. By contacting the firms directly, and in many cases, speaking directly to the manager, this investigation was able to retrieve information on the managers' state of origin; religion; languages spoken as well as their ethnic background. This data was collected during November 2012 to December 2012.

Other data that was collected for this study include: information on share prices from the Nigeria Stock Exchange; information on oil revenue from the Nigeria Bureau of Statistics and the Ministry of Finance; Information on household experiences with corruption from the Afrobarometer survey; and information on the awarding of government contracts from the Bureau of Public Procurement.

2.2 The Definitions of Corruption

Corruption is defined as the disregard of rules and regulations in the use of public office for private gain [Leite & Weidemann , 2002]. This definition covers various types of corruption: from a police officer overlooking a traffic offence by an individual in return for a bribe, to the manipulating of financial statements in order to hide the theft of funds. This dissertation focuses on the giving of informal gifts or payments from businesses to public officials in order to speed up business. This is usually termed as bribery. The Nigerian Criminal Code Act defines "official corruption" as the giving or promising to give of property or benefits on account of any act, omission, favour or disfavour of a public official in carrying out his/her duties or any governmental affairs [Laws Of The Federation Of Nigeria , 1990]. This study uses this definition of bribery since it contains the giving of informal gifts or payments to public officials in order to speed the process of doing business. Such informal payments can be made in order to receive operating licences or to quicken the pace at which goods are passed through customs.

2.3 The National Bureau Of Statistics & Economic And Financial Crimes Commission (EFCC) Business Survey On Crime, Corruption And Awareness Of The EFCC

The Economic and Financial Crimes Commission (EFCC) was established in 2004 by the Federal Government of Nigeria with the intention of revitalising the Nigerian economy. The commission has the authority to coordinate; investigate; and enforce all economic and financial legislation. The "Business Survey On Crime And Corruption, And Awareness Of The EFCC" was carried out by the Nigeria Bureau of Statistics (NBS) and the EFCC. Technical assistance was provided by the United Nations Office on Drugs and Crime (UNODC) as part of the International Crime and Corruption Business Survey (CCBS). The survey used a structured questionnaire which consisted of 10 subsections including, but not restricted to: firm characteristics; crimes; bribery and corruption; intimidation/extortion and protection money; access to the Justice System; and awareness of the EFCC. The sample of firms covered 15 economic sectors subject to the International Standards of Industrial Classifications (ISIC) Revision 3. These were: agriculture; fishing; mining and quarrying; manufacturing; electricity, gas and water; building and construction; wholesale and retail trade; hotels, restaurants and tourism; transport, storage and communication; financial intermediation; real estate, renting and business activities; public administration and defence; education; health and social work; and other community, social and personal services.

The sample frame used for selection of establishments came from the frame of establishments for: the Economic Survey and Census Division of NBS; the National Quick Employment Generation Survey (NQEGS); and the NBS/CBN/NCC Collaborative Economic Survey. Selection was based on the number of employees (10 and above); contribution of sector to GDP; and prior performance of the sector in the economy. Out of 2,775 firms that were selected for the survey, 2,215 (29.8%) were sucessfully contacted; and results from 2,110 (95.3%) of these firms were useable.

The NBS data includes a series of variables representing operations performed by companies that required them to meet with with public officials. These are listed in Table 2.1 and include: clearing goods through customs; obtaining road worthy certificates; procurement of goods and services from the government; obtaining business licenses and permits; getting clearance for environmental or sanitary regulations; obtaining residence and work permits; registering a vehicle; being involved with police investigations; having committed traffic offences; and being in contact with the courts. These variables are used in Chapter 3 as as measures of the ways in which public officers might exercise control over the firms operations. The data also includes information on whether or not a bribe was paid for each of these activities. This allows for an analysis of the the propensity to bribe conditional on a specific type of contact with a public official; for many different types of contact with public officials. This data allows for an analysis of differentiated public officers; and heterogeneity in the effect of meeting with them, respectively. Looking at different potential bribe attracting activities allows for a distinguishing of which activities are most likely to attract bribes, and which are less so.

The NBS data also has information on the different ways in which bribery takes place. That is, whether a firm offers an informal payment to a public officer, or whether a public official demands a payment from a firm in order to carry out some special favour. With this information it should be easier to determine the direction of causality with regards to the payment of bribes; it should be clear whether the market is created by public officials wanting more money; or by private firms wanting to bypass or speed up regulations[2]. The data also includes these questions for foreign public officials, which allows one to determine who is more susceptible to engage in corruption with Nigerian firms: foreign public officials or Nigerian ones[3].

The survey also included the street address of each company. This study

[2]It should be noted that there is the potential for reverse causality in this framework. I.e. a firm which has been previously demanded for a bribe by an official might tend to offer a bribe more often. Also, firms being quick to offer bribes might make public officials more likely to demand them from all firms.

[3]Comparisons of the behaviour of public officials from different countries is performed by Fisman & Miguel [2007] and Lambsdorff [1999, 2007]

Table 2.1: Data Description - NBS Data

Category	Variable Name	Definition	Measurement
operation performed by companies	conda	clearing goods through customs	Dummy Variable (0;1)
	condb	obtaining road worthy certificates	
	condc	procurement of goods and services from government	
	condd	obtainig business licenses and permits	
	conde	procurement of goods and services from private companies	
	condf	getting clearance for environmental or sanitary regulations	
	condg	residence and work permits	
	condh	vehicle registrations	
	condi	police investigations	
	condj	traffic offences	
	condk	contact with the court	
(company experience in corruption) bribe being paid in performing business operation	bribea	clearing goods through customs	Dummy Variable (0;1)
	bribeb	obtaining road worth certificates	
	bribec	procurement of goods and services for government	
	bribed	obtaining business licenses and permits	
	bribee	procurement of goods and services for private companies	
	bribef	getting clearance for environmental or sanitary regulations	
	bribeg	residence and work permits	
	bribeh	vehicle regulations	
	bribei	police investigations	
	bribej	traffic offences	
	bribek	contact with the court	

matches these addresses to their respective local government (LG) area using LG information from the Independent National Electoral Commission [INEC , 2010] and the City Population database [Brinkhoff , 2013]. The INEC [2010] database contains information on all the political wards within Nigeria: matched to their respective local government area; which in turn are matched to their respective geo-political state. The Brinkhoff [2013] database contains a geographical mapping of similar information; it shows the physical location of each LGA within each state in Nigeria. This information is used in the fourth main chapter (Chapter 6) to compare the ethnicities of the firm managers to that of the local government officials to see if sharing a common ethnicity has any effect on the propensity to pay a bribe (Section 2.6 has more details concerning the primary data collection).

2.4 The World Bank Enterprise Survey

2.4.1 Scope And Coverage

The Enterprise Survey in Nigeria was made up of a series of structured, face-to-face interviews with the owners and senior managers of a sample of 5,544 companies across the 36 states and 1 Federal Capital Territory in Nigeria; it was also conducted across most sectors of industrial activity and firm sizes.

The sample was conducted following the International Standard Industrial Classification Of All Economic Activities (ISIC) revision 3.1 [United Nations , 2002]. Following this classification, the following industries were targeted: manufacturing; construction; retail and wholesale services; hotels and restaurants; transport, storage and communications; and computer and related activities [World Bank , 2007]. For companies with 5 or more full-time permanent paid employees the sample was stratified according to the following categories: {manufacturing; retail trade; other}. Firms with 4 or less full-time permanent paid employees were randomly sampled from the population, without stratification. This sample comes from a population of over 22,000 formal companies [Iarossi, Mousley & Radwan , 2009, World Bank , 2007] and over 750,000 informal businesses [Iarossi, Mousley & Radwan , 2009].

2.4.2 Sample Design: Sample Frame

The establishment frames used for selection came from: Nigeria's Federal Office Of Statistics, The Corporate Affairs Commission; various governmental departments and authorities (State Planning Commissions, Revenue And Tax Authority, Ministry of Commerce And Industry, Exporting Zone Authority, Investment Promotion Agency); the Manufacturers Association Of Nigeria; and the Nigerian Association of Small And Medium Enterprises [World Bank , 2007].

To check the reliability of the data in the establishment frames, the name, contact details, sector/activity, and number of employees were verified by phoning a subset of businesses. 20% of formal manufacturing firms and 5% of formal non-manufacturing firms were checked during this process. During the survey process, the establishment frames were updated as new information was received concerning businesses that had closed or had changed their operations/status, and by doing so went out of the scope of the study.

2.4.3 Sample Design: Selection Procedure

The sampling strategy was designed to ensure that statistically robust analyses could be conducted with a level of precision of 7.5% for a 90% confidence interval for estimates of population proportions at the industry level. The Survey was stratified by: industry sector; firm size (number of employees); and geographic location. Stratification by firm-size was performed by dividing the economy into three groups: small firms (consisting of 5-19 employees); medium sized firms (consisting of 20-99 employees); and large firms (consisting of 100 or more employees). The stratification by geographic location was conducted to represent the distribution of non-agricultural economic activity in Nigeria. In the current case, this meant including the main urban areas of the country. For each region, the capital of each state was chosen as the main centre of population around which non-agricultural activity is clustered.

Stratification by sector was done by singling out 9 main manufacturing industries and grouping the remaining manufacturing industries together into an additional stratum. Retail and the rest of the non-agricultural economy filled up the rest of the strata with the rest of the economy being composed of the 3 main non-agricultural sectors and the rest grouped into another strata. In total, this added up to 15 different sectors of activity. In order to maintain comparability with previous studies, food & beverages (ISIC 15) and garment (ISIC 18) manufacturers were included in the sample. The other main industries were chosen based on the criteria: employment; number of firms; and contribution to GDP.

2.4.4 Gauging The Level Of Bribery

The wording on the questions concerning bribery and tax evasion in the Enterprise Survey are indirect. Rather than asking how much a firm pays in bribes the surveys query how much similar firms pay. This method of asking indirectly is thought to generate more candor on the part of the interviewees about the bribes of their own firm. This is because the indirect nature of the questions deflects any guilt away from the firm.

This method of asking indirectly is thought to be useful in gauging the level of bribe paid by the firm due to the false consensus effect [Ross, Greene & House , 1977]. This is the phenomenon that people who engage in socially undesirable behaviour are more likely to overestimate the prevalence of that behaviour amongst their peers. In this case, firms that pay bribes are more likely to say that similar firms pay bribes. This helps the current study to interpret the answer to indirect questions as being representative of the firm itself.

The indirect form of posing questions is similar to that used in the 1998 Enterprise Survey for Uganda [Reinikka & Svensson , 2003]; the Business Environment And Enterprise Performance Survey (BEEPS) on the transition countries [Hellman, Jones & Kaufmann , 2000]; and the North-African Survey of Firms conducted by Université Paris 1 [Delavallade , 2011]. This allows for a higher level of comparability between different economies than would otherwise be the case.

Further support for the validity of the data concerning informal payments is provided by the results of similar surveys carried out within Nigeria and in other countries. In a survey of businesses in Romania [Azfar & Murrell , 2009], there is no difference in the admitting of bribery between firms asked about their own businesses and those asked about “businesses like yours”. This lends favour to the argument that no less information is retrieved by asking bribery related questions indirectly.

Firm Reports About The Level Of Bribery The data from the world bank enterprise survey for Nigeria gave firm managers a choice of how to respond to the questions asking for amounts paid in bribes. Firms could choose to report bribe payments: as a percentage of sales; or in Naira currency. In the dataset used, 52% of firms said that firms similar to their own paid bribes. Out of the firms that reported a positive bribe amount; 78% reported this as a percentage of sales; and 22% reported this in Naira.

The dataset includes information on the total value of annual sales, this allows one to convert the bribes that were reported as a percentage of sales into their Naira equivalent, and vice versa. Mean comparison tests show that firms who reported bribes as a percentage of sales report significantly higher values of bribe payments; bribe payments per employee; profit and amount of moveable capital compared to the firms that reported positive bribes in terms of Naira amount. Those who reported in Naira value declared a larger proportion of their sales for tax purposes and also tended to engage in international importing and exporting more than the firms who reported in terms of sales. There were no significant differences in the number of employees; capital per worker; or receipt of government provided services between the two groups. The median bribe amounts for firms reporting as a percentage of sales and in Naira were ₦390,000 and ₦50,000, respectively.

The disparity between the reports of the firms reporting as a percentage of sales and in Naira, respectively, is not unique to the Nigerian dataset. Clarke [2011] documents results from Angola; Botswana; Burundi; the Democratic Republic Of Congo; the Gambia; Ghana; Guinea-Bissau; Guinea Conakry; Kenya; Mauritania; Namibia; Rwanda; Swaziland; Tanzania;and Uganda. Results from all of these countries are the same as the result from the dataset used in this dissertation: firms that report bribes as a percentage of sales tend to report higher amounts (when this is converted to a currency figure) than firms who report in terms of the local currency. Clarke [2011] gives possible reasons for this phenomenon, such as an imporvement in accuracy when firms report bribes a percentage of sales [Iarossi , 2006]; alternatively, firms might be less accurate when reporting smaller bribes as a percentage of sales. Clarke [2011] notes that the differences in reports in his dataset does not come from outliers biasing the averages or from the possibility that the two groups of firms are different. Despite this, the current study notes that in the Nigerian dataset, firms who reported as a percentage of sales have, on average, a larger amount of moveable capital than firms who reported their bribes in Naira currency. The Central Bank of Nigeria's (CBN) industrial policy on retail cash collection and lodgement places a limit of ₦500,000 and ₦3,000,000 on cash withdrawals and deposits by individuals and companies, respectively (Before April 2012, these limits were ₦150,000 ₦1 million, respectively). Transactions made above this limit are charged and are subject to investigation by the bank. For withdrawals, individuals (companies) are charged 3% (5%) for every amount withdrawn above the limit. For cash deposits, individuals (companies) are charged

2% (3%) for amounts deposited above the limits (Before April 2012, these charges were previously between 10% and 20%).

For companies wishing to pay a bribe much greater than this limit, and for public officials seeking to collect an amount above this limit, this regulation creates an incentive to use alternative means to make a corrupt transaction. Such a transaction might occur through the use of moveable capital goods and other assets. This might explain why firms that report bribes as a percentage of sales have a higher amount of moveable capital than the firms who report bribes in Naira. Also, the value of moveable capital might be easier to refer to as a percentage of sales rather than in Naira amount since it is probably more closely related to the amount of sales than it is to its (Naira) market value.

2.4.5 Randomised Response Questions

The randomised response (RR) questions in the survey were asked using a variant of the forced response technique that is described in section 4.2. Interviewees were given the following instructions:

"Please toss the coin handed to you by the enumerator before each question is posed without letting him/her see the results. Always answer YES if the coin comes up HEADS. Answer the question TRUTHFULLY if the coin comes up TAILS (i.e. answer YES if you have done this behavior; Answer NO if you have never done this behavior)."

In this case the probability that the interviewee is required to answer the sensitive question truthfully is 1/2, so $p = 0.5$. The probability of them having to answer "yes" regardless of the truth is 1/2, so $\theta = 0.5$. Therefore, the probability that they will be forced to answer "no" regardless of the truth is zero, $1-p-\theta = 0$. In this set-up, the respondent is only required to say "no" to a question if he/she is innocent *and* he/she receives a tails on the coin flip for that question.

The RR questions used in the analysis are:

1. Have you ever paid less in personal taxes than you should have under the law?

2. Have you ever paid less in business taxes than you should have under the law?

3. Have you ever made a misstatement on a job application?

4. Have you ever used the office telephone for personal businesses?

5. Have you ever inappropriately promoted an employee for personal reasons?

6. Have you ever deliberately not given your suppliers or clients what was due to them?

7. Have you ever lied in your self-interest?

8. Have you ever inappropriately hired a staff member for personal reasons?

9. Have you ever been purposely late for work?

10. Have you ever unfairly dismissed an employee for personal reasons?

The respondent was allowed to fill in the appropriate answers (yes/no) himself/herself. Questions 4,7 and 9 are deemed to be less sensitive [Azfar & Murrell , 2009]. The behaviour in these three questions are not illegal according to company law whereas the behaviour concerning the other 7 questions are against the law.

2.4.6 Indirect Questions

The indirect questions used in the questionnaire attempt to get information from the interviewees by: adjusting the level of detail required in the answers (e.g. by using categories instead of reported number of occurrences); using longer questions (which can generate more accurate answers when asking sensitive topics about behaviour [Iarossi , 2006]); explaining that the behaviour in question is common; following a set of warm-up questions (multiple questions concerning the effect of corruption on business operations are placed in different subsections of the questionnaire [Warwick & Lininger , 1975, Svensson , 2003]); being asked towards the middle/end of the questionnaire, by which time the questioner presumably had established credibility and trust with the interviewee; being placed amongst a series of questions about time spent dealing with government regulations, potentially being less sensitive and more appropriate in such a location; and being phrased indirectly in order to avoid implicating the interviewee of any wrongdoing. The survey was carried out by a non public-sector organisation, potentially increasing participation and attracting the confidence and trust of the participating firms. Svensson [2003] argues that awareness-raising campaigns about corruption can help to reduce the sensitivity of the topic. If so, then the presence of such campaigns within Nigeria in addition to media exposure of corruption cases might also help to make businesses more candid about the issue of corruption.

The indirect questions used in the analysis are:

1. To what extent do you agree or disagree with the following statements? (strongly disagree; tend to disagree; tend to agree; strongly agree)

- It is common for establishments in this line of business to have to pay informal payments/gifts to get things done with regard to customs, taxes, licenses, regulations, etc.
- Establishments in this line of business know in advance about how much this informal payment/gift is to get things done.

2. We've heard that establishments are sometimes required to make gifts or informal payments to public officials to "get things done" with regard to customs, taxes, licenses, regulations, services etc. On average, what percentage of total annual sales, or estimated annual value, do establishments like this one pay in informal payments/gifts to public officials for this purpose?

3. When establishments like this one do business with the government, what percentage of the contract value would typically be paid in informal payments/gifts to secure the contract?

4. What percentage of total annual sales would you estimate a typical establishment in your sector of activity reports for tax purposes?

5. What percentage of the total workforce would you estimate the typical establishment in your line of business declares for tax purposes?

Indirect questions 1,2 and 3 can provide information on the level of corruption in the business environment that might be related to RR questions 5, 8 and 10. However, of potentially greater interest is the relationship between indirect questions 4 & 5 and RR questions 1 & 2. These questions relate to tax evasion of some sort. Indirect questions 4 & 5 respectively relate to sales and workers declared for tax purposes while RR questions 1 & 2 relate to the nonpayment of personal and business taxes, respectively. This study uses these questions to test the effectiveness of indirect questioning versus RR questioning in getting interviewees to respond honestly to sensitive questions.

2.5 Centre For Law Enforcement Education Survey

To the best of the author's knowledge, no data exists for the current level of ethnic diversity in Nigeria. The 2006 National Census, which was a survey of all households in the country, did not include any questions asking people about their ethnic or linguistic group [4]. Some of the most widely used measures of ELF data back to 1972 [Taylor & Hudson , 1972]; and whilst the CIA World Factbook provides estimates of the population of some of the most populous and politically

[4]Correspondance with the National Population Commission (2010)

influential groups, these do not represent a comprehensive or accurate view of the 250+ ethnic groups contained within the nation [Central Intelligence Agency , 2013]. This lack of recent data in ethnicities poses a potential problem for the analysis of ethnicity and bribery.

In order to solve the problem of lack of data on ethnicities in Nigeria this study uses information from a representative household survey that was conducted in all 36 geo-political states and the Federal Capital Territory. The survey was conducted by the CLEEN Foundation between February and May 2010. The total sample size was 10,228, which comprised of equal percentages of males and females of 18 years and above. Data was collected via in-home face-to-face interviews using a stratified multi-stage random selection procedure. The interviews were done by a Lagos based social research company, Practical Sampling International.

The survey covered questions about patterns of crime within the country and also asked respondents for general information, including: marital status; relationship to the head of household; the state and local government area that they reside in; their religion; and their ethnicity. The question about ethnicity was asked towards the end of the survey, when the interviewer would probably have developed some rapport with the manager [Svensson , 2003]. This question received no non-response. The current investigation uses this data to generate national; state; and local government level indices of ELF for Nigeria and its sub-national states and local government areas, respectively. To the best of the author's knowledge, this is the first study that generates and uses sub-national ELF in this way.

2.6 Fieldwork And Primary Data Collection

Primary data was collected for this study, this comprises of a survey of managers from the businesses contained in the NBS/EFCC survey described in Section 2.3. All firms from the NBS/EFCC dataset were contacted. Primary data on the ethnic group; gender; religion; languages spoken; and state of origin of the managers of the respective companies in the NBS/EFCC survey is used in this study. This data was collected by contacting the companies directly. Telephone calls were made between November 2012 to December 2012 to each of the 2,110 businesses in the sample. Out of the 2,110 companies, 1,267 companies were successfully contacted, this represents 60% of the total sample. Results from all 1,267 managers were analysable. Since the NBS/EFCC survey was conducted in 2007, information was collected about the managers in 2007 and not the manager at the time of calling. Information about the ethnicity and state of origin of the managers allows this study to account for the movement of labour within the borders of Nigeria, it also removes the restriction of assigning one ethnic group to a geographic area. This study combines the data on ethnicity with the information on the State and LGA of the company to create measures of ethnic diversity at the national; state; and local government level.

2.7 Basic Summary Statistics From The ES and NBS Datasets

Summary statistics for the Enterprise Survey and the NBS Survey are shown in Tables 2.2 and 2.3, respectively. The figures in both tables refer only to the manufacturing firms of the respective datasets.

Table 2.2: Summary Statistics For The Enterprise Survey Dataset

Variable	Statistic	All Firms	Non-Bribers	Bribers
Bribe Amount (in '000 Naira)	Mean	294	0	572
	Median	8	0	264
	Std. Dev.	652	0	816
Bribe Amount per employee (in '000 Naira)	Mean	20.6	0.0	40.0
	Median	0.5	0.0	22.2
	Std. Dev.	49.6	0.0	63.3
No. of employees	Mean	17	18	16
	Median	12	13	12
	Std. Dev.	16	18	14
Profit per employee	Mean	158.7	155.3	161.8
	Median	90.0	77.8	100.1
	Std. Dev.	216.8	211.8	221.4
Capital per worker	Mean	230.2	190.9	267.2
	Median	100.0	75.0	125.0
	Std. Dev.	718.8	544.4	850.0
Index of public goods received from government (0-2)	Mean	0.29	0.25	0.32
	Median	0.00	0.00	0.00
	Std. Dev.	0.50	0.48	0.51
Percentage of sales declared for tax purposes	Mean	69.6	66.1	72.8
	Median	75.0	70.0	75.0
	Std. Dev.	27.4	28.2	26.2
N		2001	972	1029

In each column, for each variable, the table reports the mean, median and standard deviation for each (sub)sample. The last row details the number of observations for each (sub)sample.

The first column of figures in Table 2.2 shows the mean, median and standard deviation of the variables for all manufacturing firms. The sample size for the summary statistics in this table is 2,001 firms. The second column of figures relate to the manufacturing firms who reported that similar firms paid a zero bribe payment to government officials, while the last column of figures relates to firms who reported that similar firms made a positive bribe payment to goverment officials. Profit is measured by deducting the previous years operating costs from the value of sales in the previous year. Capital is calculated as the net book value of machinery, vehicles and equipment at the end of the previous year. The goods, which were enquired about, that the firm could receive from the government were: water and electricity. All figures denoting money are shown in ‘000 Naira (₦).

Table 2.3: Summary Statistics Of Bribery And Meeting With Officials - NBS Dataset

	(1)	(2)	(3)
	Bribery/ Meeting With Official		
	All Firms	Met With An Official	Bribed An Official
Met With An Official	0.37	1	1
(Mean) Average Number Of Types Of Officials Met With	1.66 (2.72)	4.43 (2.74)	4.20 (2.50)
Bribery Episode	0.14	0.36	1
(Mean)Number Of Types Of Bribery Episodes	0.44 (1.39)	1.16 (2.07)	3.20 (2.31)
Observations	331	124	45
(Mean)Value Of Bribes ('000 ₦)(₦)	0.616 (4.56)	1.65 (7.36)	3.43 (9.49)

The unit of observation is the firm. Standard deviations are in parentheses.

For Table 2.3, non-parenthsised figures are mean averages, the parenthesised figures are the standard deviations. The sample size for the summary statistics in this table is 331 firms because of the smaller sample size for the NBS survey. The first column of numbers is based on all firms; the second column is based on firms that met with at least 1 of the public officials that were asked about; the third column of numbers is based on firms that met with at least 1 public official and bribed at least 1 public official. A total of 11 types of public officials were asked about and meeting with an official was a pre-condition for paying a bribe to that official. The table describes the number of types of public officials met with and not the number of public officials met with. “Met with an official” is a dummy equal to 1 if the firm met with at least 1 official, 0 otherwise. “Bribery Episode” is a dummy equal to 1 if the firm paid at least 1 bribe, 0 otherwise.

The figures in the table show that 37% of all firms (124 firms out of 331) met with at least one type of official in the year preceding the survey. Firms met with a mean average of 1.66 types of official; this average rises to 4.43 types of official when focusing solely on firms who met with an official, i.e. ignoring the zeros. Fourteen percent of all firms bribed an official (45 firms); out of the firms that met with an official, 36% of them bribed an official. Those who paid a bribe to an official did so, on average, to 3.2 types of public official. The average bribe paid by bribing firms was approximately ₦3,400.

The relevance of this data comes from the fact that the probability of paying a bribe to a specific type public official can be conditioned on the occurrence of a meeting with that type of official. This allows for the analysis to include a two-stage model that estimates the propensity to meet with a type of public official (e.g. a police officer); and then estimate the propensity to bribe the same type of official conditional on meeting them. The nature of the business operations that the firms engage in mean that they are only able to pay a bribe if they physically meet with a public official; bribes are not paid electronically. Hence, not meeting with a police officer is a perfect predictor of not bribing a police officer. Using this data, the analysis builds on previous work that models the unconditional propensity to bribe [Svensson , 2002].

3 Factors Influencing The Propensity To Bribe And Size Of Informal Payments: Evidence From Formal Manufacturing Firms In Nigeria

Abstract

This chapter uses two unique datasets on manufacturing firms in Nigeria to examine the factors which influence who pays a bribe and how much is paid. The study finds strong evidence for the control rights hypothesis: meetings with public officials are positively associated with the incidence of bribery. Evidence is also found for the bargaining hypothesis: the amount of bribe that a firm pays will depend on its current and future ability to pay as well as its outside options. The investigation contributes to the literature by disaggregating: type of bribe; and instigator of the bribe transaction. Among the different types of informal payment, traffic offences seem to attract the most bribe demands (from public officials) and offers (from companies).

3.1 Introduction: The Payment Of Bribes When Doing Business

Firm chracteristics and the institutional framework shape the way business is done and lead to different outcomes with regards to the payment of bribes. Using data on firms in the formal manufacturing sector in Nigeria, firms are asked about the bahaviour of similar firms when coming into contact with public officials: whether they are bribers (similar firms made at least one informal payment in the 12 months preceding the survey); or non-bribers (similar firms did not make an informal payment in the 12 months preceding the survey). The distinguishing of firms by their reports of "similar firms" is a useful strategy to understand the determinants and consequences of firm-level bribery. Reports of similar firms are believed to be representative of the firm itself. The bribing behaviour adopted by firms is seen as the result of an optimising behaviour in an industrial organisation scenario and is an indicator of firm-level responses to potential constraints in the business environment. In urban Nigeria, the business environment is characterised by micro; small; medium; and large enterprises. Few firms comprise a monopoly, and the decision to allocate money towards bribe payments for the purpose of complying with regulation is likely to be influenced by the relative lack of alternative options and the profitability of the firm.

Due to the secretive and somewhat tabooed nature of bribery, it might appear impossible to collect quantitative information on bribe levels from firms [Reinikka & Svensson , 2003]. On the face of this, any effort at estimating the proportion of firms engaging in bribery would be subject to bias from underreporting. However, this chapter provides evidence for the contrary view: that, by using specially designed data collection strategies, useful information can be retrieved from questions concerning bribery. This makes it possible to derive estimates of: the proportion of firms that pay bribes; the amount of bribes that these companies are required to pay; and the distribution of the bribe payments[5].

Company-level bribery can sometimes occur when a firm engages in international trade [Lambsdorff , 1998]. An importing or exporting firm might offer an informal payment to a public official if they are handling banned cargo or have failed to declare the contents of their cargo and wish the official to overlook this. A report by the United Nations [United Nations Office On Drugs And Crime , 2010] states 23 (24) different vulnerabilities of the customs service (transportation sector) that make it more likely for a business to pay a bribe to a public official. These include a) the active presence of numerous government agencies in the ports and; b) the wide powers given to the customs officials.

[5]Despite the data collection strategies, however, there might still be cases of misreporting of bribe payments in the sample, nevertheless, this chapter argues that the strategies employed significantly reduce the bias that might be inherent in answers to sensitive questions.

Having many types of government agencies at the ports make the potential importer/exporter subject to control by any one of these agencies[6]. This can lengthen the process of clearing goods. Furthermore, the clearing process can involve duplicated procedures that send the importers from one government agency to another. These requirements also create a potential barrier to trade by delaying the clearance of goods. These factors create an incentive for the importing/exporting firm to give facilitation payments[7] (or "grease money") to speed the process of goods clearance [Ibid].

Customs officials are granted powers by the Customs and Excise Management Act (CEMA) of 1958 to delay the clearance of cargo without the adjudication of any independent arbiter. This potentially leaves the process of importing/exporting at the discretion of the customs official. In addition, any fees for container rental charges or demurrage incurred during any hold up are borne by the importer/exporter. On top of this, under the regime of destination inspection (DI) which began in 2006, importers of goods that did not receive a Clean Report Of Inspection (CRI) were given a deadline of three months to clear them, after which the goods were up for confiscation by the government. These added constraints also increase the chance of a firm providing an official with an informal gift in order to speed up clearance and avoid the added costs of delay[8].

Various policies have been introduced to address corruption. Many of these reforms, or proposals, have targetted the discretion/power of public officials in dealing with private agents [The Foreign Corrupt Practices Act , 1998, Organization For Economic Co-Operation And Development , 1997, United Nations Office On Drugs And Crime , 2005]. It has been argued that the constant evaluation of managerial and business performance can reduce administrative problems and bottlenecks in the public sector and therefore reduce the possibility of bureaucrats extracting bribes [Rose-Ackerman , 1996, World Bank , 1997, Ades & Di Tella , 1999, Otusanya , 2013]. This chapter seeks to examine the extent to which public officials are to blame for the occurrence of bribery within Nigeria. This is done by examining the causes of: the propensity to pay a bribe; the size of the bribe; the frequency of the bribe payment; and the instigator of the transaction.

The rest of this chapter is organised as follows. Section 3.2 discusses the eco-

[6]Agencies present at Nigerian ports include: Immigration, the Police, Port Quarantine, NDLEA (National Drug Law Enforcement Agency), SSS (State Security Service), NAFDAC (National Agency For Food And Drug Administration And Control), and SON (Standards Organization Of Nigeria)

[7]A facilitation payment is defined as any money given to a public official with the purpose of speeding up a routine government action [The Foreign Corrupt Practices Act , 1998] (i.e. money to speed up the doing of an act which is not illegal).

[8]The report also presents information on bribery in the transport sector. Payments ("kickbacks") for awarding construction contracts or for information on competing bids; and bribes for overlooking unfit road vehicles were listed as some of the bribe attracting activities.

nomic literature on bribery and introduces a framework to model the supply of bribes. It also describes how this study builds on the previous work. The data and variables to be included in the model are outlined in section 3.3. Section 3.4 presents the methodology to be used. Section 3.5 compares the two datasets that are used in this study. Sections 3.6 to 3.10 present the empirical results from the World Bank dataset; and the NBS dataset, respectively. Finally, section 3.11 concludes and suggests potential avenues for future research.

3.2 Literature Review And Conceptual Framework

This section reviews the literature on bribe payments at the firm level; states the contribution of this chapter to this literature; and explains the significance of this contribution. The study of corruption is riddled with questions concerning the validity of the data [Kaufmann, Kraay & Mastruzzi , 2006]. This is due to the socially undesirable and sometimes illegal nature of the act, leading to the under-reporting of bribery. This makes any measure of bribery prone to measurement error due to missreporting. Therefore, any estimate of the prevalence of bribery among a sample of firms is likely to be downward biased.

Another potential source of error when dealing with firm level measures of bribery is sampling error [Ibid.]. Some industries and regions might be more prone to bribery than others, therefore, any inference based on these industries alone might give a biased estimate of the level of bribery. Furthermore, when dealing with an environment where informal payments are sometimes required to obtain operating licences and electrical connections, potential firms might be unable to enter a particular market because of this, i.e. the demand for informal payments might act as a barrier to entry for some industries. This is another potential source of sampling error because any estimate of the proportion of bribing firms will necessarily exclude firms which are not operating due to an inability or unwillingness to bribe (or both). Nevertheless, the stratified sampling and indirect questioning methods allow for reasonable inferences to be made about the sample which can be extrapolated to the population even in the presence of sample selection.

Previous studies on the factors influencing corruption have used a cross-section of countries. This body of literature has uncovered the following factors to be associated with lower levels of corruption: a greater proportion of women in parliament, the civil service, and the labour force [Swamy, Knack, Lee & Azfar , 2001]; higher relative[9] wages for public officials [Van Rijckeghem & Weder , 1997]; and a common law legal system [Treisman , 2000]. Many of the determinants do not vary below the country level; and many are not subject to change from specific govern-

[9]Ratio of Civil Service wages relative to (private sector) manufacturing wages

ment policies. The data used in this paper allows for the study of corruption with variables that vary across firms and types of public official. This contributes to the economic literature on the study of corruption and provides practical implications for policy interventions.

Framework For Modelling The Supply Of Bribes This section describes the theoretical framework that this chapter uses to analyse the factors which affect the supply of bribes amongst manufacturing firms in Nigeria. The subsequent empirical analysis is based upon this framework. Svensson [2002] introduces a model to understand the factors which influence bribery: both the occurrence of bribery and the amount paid as a bribe. This chapter uses a variation of this model, developed by the author, to examine the supply of informal payments amongst firms in Nigeria. The model considers an economy with a large number of private firms and bureaucrats (public officials), with each company being under the reach of one bureaucrat. Officials act as profit maximisers, therefore, they seek to maximise the amount of bribes that they can extract from each firm, subject to the constraints that: they might get caught and punished; or the firm might exit the market and no longer remain under their control, therefore being no longer able to pay them a bribe.

Public officials have the ability to demand payments from firms as much or as little as they please. This ability stems from their "control rights" over the firm, that is, the extent to which the firm is bound by regulation to approach the public official for public goods/services required in the course of business. These include: business licences; customs clearances; power grid connections; and special operating permits. These control rights mean that the public official is able to potentially affect the business decisions and cash flow of the company. For example, a firm that cannot clear a good through customs, due to a bribe-demanding official, faces an income constraint due to a restriction in the market that it can supply to.

The level of control rights that the official has depends on the extent to which its decisions affect the firms' outcomes. I.e. low control rights mean that the firm can refuse to pay a bribe without major consequences on its output decisions or sales, whilst a bureaucrat with a high degree of control is more able to influence the decisions of the firm by demanding a bribe. A bureaucrat might be said to have a low level of control rights over firms in the informal sector due to their activities being relatively unseen and/or unregulated, and a high degree of control rights over companies in receipt of government services, such as electricity or water supply, in order to run their business. Firms who import or export directly can also be said to be under an official with high control rights due to the clearance of goods depending on the discretion of a customs official.

A relatively simple case of two sectors, $S = \{s_1, s_2\}$, is proposed. Firms in sector s_1 can refuse to pay bribes without inconvenience, whilst firms in sector s_2 must pay bribes if demanded. A public official who interacts with a firm in sector s_2 will demand a bribe if the expected benefit from receiving a bribe is more than the expected cost. That is, $b - pcb > 0$, where b is the amount of bribe; p is the probability of being caught and punished; c is the cost of being caught and punished[10].

Moving to the demand for bribes, the distribution of c is allowed to be uniformly vary between $[0, \bar{c}]$ (as in [Erard & Feinstein , 1994, Ades & Di Tella , 1999]), therefore allowing for the existence of both honest and corrupt officials. Furthermore, all bureaucrats face an exogenously given probability $1 - f$ of being fired, allowing for the uncertainty of their post. The wage rate in each sector is normalised to zero and at time $t = 0$ officers must choose which sector to work in. Each sector employs half of the population of officials. An official who is indifferent between working in either sector is randomly assigned to a sector that has room for employment.

In equilibrium, all bureaucrats with cost $c \leq 1/p$ choose to work in sector s_2, while those with cost $c > 1/p$ are randomly assigned to the remaining posts. The probability g that a randomly selected official in sector s_2 will ask for an informal payment is

$$g = \begin{cases} 1 - 1/(2pc) & if \quad pc > 2 \\ 1 & otherwise \end{cases} \tag{3.2.0.1}$$

.

Public officials who wish to demand bribes will choose to work for government sectors where control rights are high.

A randomly picked company is required to pay a bribe with probability:

$$d(i) = \rho(i \in s_2) * g \tag{3.2.0.2}$$

, where $\rho(i \in s_2)$ is the probability that firm i operates in sector s_2.

Given that a firm interacts with a corrupt bureaucrat in a given period t, it is possible to solve the probability of the firm being demanded a bribe in each future period evaluated from time t. This probability function is defined as: $d_{t+n} = f^n + (1 - f^n)g = f^n - f^n g + g$. Hence, if a firm has already been matched with a corrupt official in period t, then the probability of the firm being asked to pay a bribe in period $t + n$ will be the summation of: the probability that the official is not fired by period $t + n$; and the probability that the next bureaucrat will ask for a bribe in period $t + n$ (if the initial official had been fired by period $t + n$).

[10] This assumes the the cost of being caught is proportional to the amount of bribe paid.

Each firm has the objective of maximising the present value of discounted cash flows (profits minus informal payments). Firms in sector s_2 have capital k and firm specific knowledge τ of production. τ is distributed according to a known distribution function $T(\cdot)$. The share of capital that can be resold and/or reinvested is denoted by κ^i. This measures the degree to which capital is (not) sunk (a low κ^i implies a larger share of sunk capital). At time 0 each firm has the choice of investing in either sector s_1 or s_2 but not both.

Companies in sectors 1 and 2 produce goods x_1 and x_2, respectively. Both goods are traded internationally. The country is a price taker and so market prices in sectors 1 and 2 are given by 1 and δ, respectively. Production technologies are given by ${x^i}_1 = f(k^i, l^i)$ and ${x^i}_2 = f(k^i, l^i, \tau^i)$ where l is labour; and production is increasing with τ. The supply of labour is unlimited at the given wage rate w. The price of the second good is uncertain, i.e. $\delta_t \in [\underline{\delta}, \bar{\delta}]$ is a random variable that is independently & identically distributed over time.

Profit is sector s_1 is given by:

$$\pi(k, l(w)|s_1) = f(k, l(w)) - wl(w).$$

Profit in the second sector s_2 at time t is given by:

$$\pi(k, l(w/\delta_t); \tau, \delta_t|s_2) = \delta_t f(k, l(w/\delta_t); \tau) - wl(w/\delta_t)^{11}.$$

In a situation with no bribery, firm i has the value function:

$$\mathbf{V_t}(k|S) = \mathbf{E_t}(\sum_{n=1}^{\infty} \beta^{n-1} \pi(k, \cdot|S)) \quad for \quad S = \{s_1, s_2\}.$$

where $\mathbf{E_t}(\cdot)$ represents the expectation of the term enclosed in brackets conditional on the level of information at time t.

Once requested for a bribe, firms in sector s_2 can choose to either pay the bribe or exit the industry. Firms will choose to exit the market if the gain from exiting (the alternative return from using their capital elsewhere) is greater than the gain from paying the bribe and staying in the market (current and future net profits).

$$\sum_{n-1}^{\infty} \beta^{n-1} \pi(\kappa k, \cdot|s_1) \geq \pi(k, \tau_t, \cdot|s_2) - b(\tau_t) + E_t \sum_{n-1}^{\infty} \beta[\pi(k, \tau_{t+n}, \cdot|s_2) - d_{t+n}(s_2) b(\tau_{t+n})] \tag{3.2.0.3}$$

where $b(\tau_t)$ is the amount of bribe paid in period t as a function of τ_t. The first term of this expression is the discounted profit the firm would receive if it

[11] Firm subscripts have been dropped. The labour demand function is given by $l(w/\delta_t)$. The first order condition: $\delta_t f_l(k, l; \tau) - w = 0$ means that firms adjust their number of employees to make the marginal product of labour equal to the real wage rate.

sold and reinvested its capital in sector s_1 during period 1. The first two terms on the right hand side denote net profit when faced with a bribe demanding public official. The third term on the right hand side represents discounted future net profits.

Firms cannot borrow in order to pay bribes so in each period net profit must be non-negative:

$$\pi(k, \tau_t, \cdot|s_2) - b(\tau_t) \geq 0 \quad \forall \quad t. \tag{3.2.0.4}$$

The amount of bribery in equilibrium is determined as follows. Given (3.2.0.4), the bribe maximising official will demand payments so that (3.2.0.3) becomes an equality. Also, expected future net profit $E_t\pi(\cdot)$ is constant and independent of the price of x_2 because bribes absorb any excess over reservation level. Therefore, each period, the exit constraint (3.2.0.3) is identical except for the first term on the right hand side, current profit. Rearranging this expression with an equality constraint gives:

$$b(\tau_t) = \pi(k, \tau_t, \cdot|s_2) + E_t \sum_{n-1}^{\infty} \beta[\pi(k, \tau_{t+n}, \cdot|s_2) - d_{t+n}(s_2)b(\tau_{t+n})] - \sum_{n-1}^{\infty} \beta^{n-1}\pi(\kappa k, \cdot|s_1) \tag{3.2.0.5}$$

This expression has $b(\tau)$ on both the left hand side and the right hand side. It therefore maps $b(\tau)$ onto itself. The fixed point value of $b(\tau)$ is determined by:

$$b^*i(\tau_t) = \pi(k, \tau_t, \cdot|s_2) + d''(1+d')^-1\bar{\pi}(k, l^i, \cdot|s_2) - (1-\beta)(1+d')\pi(\kappa k, \cdot|s_1)^{12} \tag{3.2.0.6}$$

Equation (3.2.0.6) states that the amount that a business will need to give in informal payments is positively related to current and expected future profits, and negatively related to the profit gained from reallocating capital elsewhere ($\pi(\kappa k)$). I.e. higher current or future profits weakens the firms "bargaining power"[13] whilst lower sunk costs strengthen its bargaining position.

Equation (3.2.0.6) also implies that the equilibrium bribe amount $b(\tau_t)$ is negatively related to g but that the expected bribe amount, $d * b(\tau_t)$ is positively related to g. That is, a lower probability of a public official demanding a bribe will mean a lower expected bribe to be paid but a higher equilibrium bribe to be paid (when a firm interacts with a corrupt official). The reason for this comes from the explanation about bargaining power. A lower probability of bribe demands

[12] Where $\bar{\pi}(k, \cdot|s_2) \equiv E_t\pi(k, \tau_{t+n}, \cdot|s_2) \quad for \quad all \quad n \geq 1$, and $d' \equiv \beta(f(1-\beta) + g(1-f)(1-\beta)^{-1}(1-f\beta)^{-1})$, and $d'' \equiv \beta(1-f)(1-g)(1-\beta)^{-1}(1-f\beta)^{-1}$.

[13] The ability to bargain for a lower bribe amount.

will increase expected future profits, these will weaken the bargaining power of the firm, so the bureaucrat can demand a higher payment[14].

Solving equation (3.2.0.4) to find the conditions under which this constraint holds gives:

$$1 - \frac{(1 - f\beta)\pi(\kappa k, \cdot|s_1)}{\beta(1 - f)\bar{\pi}(k, l^i, \cdot|s_2)} \leq g. \tag{3.2.0.7}$$

This means that a sufficiently high probability of a bribe being demanded, g, ensures that the equilibrium bribe amount is less than profit [Shleifer & Vishny , 1993]. Equations (3.2.0.2) and (3.2.0.6) present two formulas with which to estimate the behaviour of bribe payments amongst firms in Nigeria. Whether or not a firm pays a bribe depends on where it chooses to locate (industry-wise) and the expected cost of the bureaucrat being punished for engaging in corruption. In the case that a firm meets with a public official and pays a bribe, the amount paid will depend on company specific characteristics. These include: current profits; expected future profits; and the sunk cost component of capital. The current study extends the empirical specification of this model by including more variables to measure the control-rights hypothesis. This allows for a broader view of the things that potentially determine whether or not a firm pays a bribe.

Certain assumptions are inherent in this analysis. In the real world public officials do not have full information about firms. Price shocks, τ, and profits, π, are not observable. This potentially gives the firm a stronger bargaining position than is shown in the model. Nevertheless, introducing asymmetric information into the model does not affect the qualitative results [Svensson , 2002]. In the model, each firm is matched to only one bureaucrat. If the firm can meet with more than one bureaucrat, each providing the same service/licence, then the results will not be qualitatively affected if both officials choose to collude. If, however, they choose to compete for bribes (i.e. outbid each other by demanding a lower bribe than the other bureaucrat(s)) then the refusal power of the firm might increase [Shleifer & Vishny , 1993]. In the case where a firm needs two complimentary licences in order to function, each of which is provided by a different public body, providers of these licences will tend to act independently: different ministries and agencies set their own bribe level to maximise their individual bribe revenue rather than the combined revenue of the bribe collectors. This scenario is prevalent in Nigeria (amongst other African countries), India and post-Communist Russia [Ibid.]. In order to start a business in Nigeria a company is required to register with the Corporate Affairs Commission; the Federal Board of Inland Revenue; the Ministry of Finance; and the State Tax Office. This is in addition to receiving inspection

[14] A fall in the probability of meeting a corrupt official, d, cannot be outweighed by an increase in bribe payment $b(\tau_t)$. This is because this would suggest that a lower d would increase $b(\tau_t)$ and therefore result in lower expected future net profits and lower $b(\tau_t)$, a contradiction.

from the local government and paying fees at a designated bank [World Bank , 2011]. These facts tend to suggest that the model adequately describes the nature of bribery in Nigeria.

Another assumption in this model is that profits are not determined by the amount of bribes paid, and that entry & exit into the market do not affect equilibrium profits. Despite this, there is evidence to suggest that these restrictions are valid in this model. Firstly, the sample of firms used is populated with mainly small-sized firms (a median of 12 employees in the enterprise data). Research suggests that the regulatory system is not captured by such firms but by bigger companies with a relatively large amount of political power [Delavallade , 2011]. Secondly, the nature of uncertainty surrounding a public officials tenure in office [Thomas, M. , 1999] suggests that they heavily discount the future. Therefore, any graft scheme that aims to maximise bribe revenue by controlling entry and exit into a market does not seem plausible because bureaucrats will try to acquire as much bribe income as they can as quickly as possible in case they are soon fired or transferred to a new position. This seems to occur within the Nigerian Police Force, reports [Human Rights Watch , 2010] detail police officers being required to pass on a portion of their collected bribes up the chain of command to their superior officers who pay their superior officers in turn. Officers sometimes bribe their superior officers in order to be assigned to positions which have a relatively better opportunity to demand bribes from civilians. In return, superior officers sometimes set monetary targets to be achieved by the lower-rank officers in these positions. Individuals that fail to collect (and pass on) a minimum amount of bribes can be transferred to another position with a lesser chance of collecting (an receiving) bribes. Despite these arguments in favour of the conceptual framework there is literature that predicts a positive relationship between bribes and profits [Bliss & Di Tella , 1997]. Section 3.6 attempts to empirically resolve the issue of reverse causality issue by instrumenting for profits. The next section discusses the data used to estimate the supply of informal payments.

Empirical Work On The Supply Of Bribes Svensson [2003] uses data on Ugandan firms to describe the factors driving: the giving of informal payments; and the size of these gifts. The first of these is linked to the level of required contact that the firm has with public officials (the control rights hypothesis). On the other hand, the amount of bribe that is paid is linked to the firm's ability to pay and refusal power (bargaining power). Whilst the results are statistically significant, the analysis suffers from a relatively small sample size (230). Out of 230 companies, only 176 answered the bribery-related question. Out of these firms only 143 reported a positive bribe, therefore, the analysis on bargaining rights was restricted to this sub-sample. Also, the estimation included outliers which, in a smaller sample, are more likely to bias the results. The current chapter builds on this work by using a much larger sample size, this gives more power to the study and decreases the probability that the results are driven by outliers.

This study also builds on the previous literature by increasing the number of variables used as measures of meetings with government officials. In addition to the receiving of public services; whether or not the company engages in international trade; and the amount of tax that is declared, the analysis includes a dummy indicating whether the government was the principal purchaser of the establishment's output; and an index of tax exemptions (section 3.4). Adding to the number of variables that fall under the control rights hypothesis helps to strengthen the results. If all variables (original and added) return significant coefficients and their expected signs, then this will give more weight to the argument that increased contact with public officials (who have discretion over the granting of services and the receiving of bribes) increases the chance of a company paying a bribe.

[Delavallade , 2011] explores the link between tax evasion and bribery. This adds to previous work which finds bribery and tax evasion to be substitutes (firms leave the formal sector in order to avoid bribes; and firms in the formal sector are less able to escape being targeted for bribes) [Rand & Tarp , 2010] and work that suggests that their relationship might be complementary [Johnson, Kaufmann, McMillan & Woodruff et al. , 2000] (companies might pay bribes to tax officials in order to be allowed to hide their sales). Delavallade finds a threshold level of tax evasion[15], before which the complemantary relationship persists and after which the substitutary effect dominates. One potential limitation of the Delavallade [2011] paper is that the measure of bribery is a categorical ordered response to the question of whether similar firms have to give informal gifts to public officials in order to work. The responses are: never; seldom; sometimes; often; mostly; and always. Whilst this gives a measure of the frequency of bribery, it does not measure the level of bribes. Previous work [Méndez & Sepúlveda , 2010] shows that an estimate of the prevalence of bribery can change depending on whether one uses the absolute frequency of bribe paying transactions; the frequency of bribe paying transactions as a share of the total number of transactions; or the aggregate amount of money paid in bribes. The current study adds to the literature by complementing the categorical measure of bribe frequency with a currency measure denoting the actual bribe amounts paid (in '000 Naira) by bribing firms. This allows an analysis of not only the factors contributing to the propensity to bribe and frequency of bribe payment, but also the factors influencing the bribe amount. In doing this, the current chapter will use two[16] out of the three bribe measures

[15] The threshold occurs where the firm hides 55% of sales.

[16] The analysis in chapter 6 uses all three measures defined in [Méndez & Sepúlveda , 2010] and adds a fourth. The results of these estimations are shown in the Appendix Section A.4.3.

stated in Méndez & Sepúlveda [2010], the previous studies listed [Svensson , 2003, Delavallade , 2011, Rand & Tarp , 2010] use only one of the three measures. In doing this, this paper will provide a more comprehensive discussion of the nature of bribery and the factors determining the prevalence; magnitude; and frequency of bribe payments amongst manufacturing firms in Nigeria.

Adding bribe amounts to the analysis will also help to decrease potential response bias that can occur when a question is asked that requires the opinion of the interviewee. Two firms, both identical in observable features, might have to pay the same amount of bribe the same number of times per month. Nevertheless, the firms might respond differently to the bribery related question used in Delavallade [2011]. This bias hinders an unbiased comparison between firms (whether similar or not), however, the use of bribe amounts might help to reduce this bias.

Unbundling the concept of corruption can help to make progress with finding out facts relating to it (Mishra, 2009)[17]. Fan, Lin & Treisman [2010] showed that different types of corruption can have different effects on an economy and economic agents might settle for certain types of corruption as a second best option due to the worse effects of other types. This variability in the nature of corruption can also be applied to bribery: different types of bribery might have different drivers. Due to the variety of activities which may involve a bribe payment, having to pay bribes "to be able to work" is open to different interpretations which might lead to a bias in the responses of firms. The response to the question might depend on what work the company was involved in. Paying bribes in order to clear goods with customs would be different from paying bribes in order to bypass punishment for traffic offences, yet both activities might be what different firms mean by being able to work[18]. A bribe to a customs official might also be much higher than that paid to a traffic police officer. This chapter adds to the adds to the body of knowledge by unbundling the concept of bribery. The prevalence and reasons for different types of bribery are laid out. This allows for an understanding as to which activities attract the most graft. The data used in the rest of this chapter is described in the next subsection. The next subsection also outlines the construction of variables and their use in the analysis.

[17] Correspondance with Mishra.

[18] The survey contains two other bribery related questions, one of which asks whether similar firms had to make informal payments to start their activity, and the other asks whether they have to make informal payments to accelerate administrative procedures.

3.3 Data And Variables Measuring The Determinants And Scale Of Bribery

Data for this chapter is taken primarily from business surveys within Nigeria. These are: the Nigeria Bureau Of Statistics (NBS) and the Economic And Financial Crimes Commission (EFCC) Business Survey On Crime And Corruption And Awareness Of The EFCC; and the Enterprise Survey For Nigeria. Both NBS and Enterprise surveys include businesses from all 36 states of Nigeria as well as the Federal Capital Territory. The datasets also cover the manufacturing sector, which is the focus of this chapter. Both surveys also ask questions on bribery, and more specifically, the relative frequency with which bribes are paid and the amount paid.

Variables Used This section describes the variables used in the forthcoming analysis. The data is described in Tables 3.1 and 3.2 for the first (World Bank) data set; and Tables 3.3; and 3.4 for the second (NBS) set of data. All currency figures are measured in '000 Naira. In general, the variables represent either the control rights or the bargaining mechanisms, respectively.

Current profit is measured as the value of total sales minus operating costs and interest payments[19] [20]. Capital stock is calculated as the resale value of machinery, vehicles and equipment. The sunk cost component of capital stock represents the ability to reallocate capital elsewhere. It is estimated using data on the resale and replacement value of capital stock[21]. The resale value is how much it would cost to purchase the equipment in its condition at the time. The replacement value of capital is defined as the cost of replacing all machinery, vehicles and equipment with new ones. The ratio of the resale value to the replacement value of capital represents capital mobility and the extent to which it has physically depreciated: older capital is less productive and has a lower resale to replacement value. To measure capital mobility, this chapter follows the method used in Svensson [2003] and regresses the resale/replacement ratio on a constant and the average age of capital stock. The residual of the regression represents that amount of the resale/replacement ratio that is not captured by the age of machinery, vehicles and

[19] Operating costs are measured as the sum of the following costs: raw materials and intermediate goods; labour (including wages, salaries and bonus/social payments); depreciation; rent of land/buildings, equipment and furniture; electricity; fuel; water; transportation (excluding fuel) and communication services.

[20] Loan providing financial institutions include: private commercial banks; state-owned banks and government agencies; non-bank financial instutions; and informal sources of credit.

[21] This is similar to Ramey & Shapiro [2001], who use the resale price to reflated initial cost of capital to measure the cost of reallocating capital across firms and sectors.

Table 3.1: Data Description

Variable Name	Definition
bribereport	Dummy=1 if bribe reported in Naira, 0 if as a percentage of sales
bribe_naira	Reported bribe payment (in '000 Naira)
bribe_dummy	Dummy=1 if positive bribe reported, 0 otherwise
bribe_gov_contract	% of gov contract paid in bribe to secure the contract
tax_exemption	Index (0-4) of tax exemptions. Import, profit, VAT, and export taxes.
request_government_services	Index (0-6) of request for services in last 2 years: phone/electricity/water connection; import/operating license; construction permit
tax_visit	Dummy=1 if firm was visited, inspected by or required to meet tax officials in last year
tax_percentage	Percentage of sales reported for tax purposes
employee_total	Total number of full-time (permanent and seasonal/temporary) and part-time workers
size_1	size==small (5-19 employees)
size_2	size==medium (20-99 employees)
size_3	size==large (100 employees and more)
foreign	Dummy=1 if foreign ownership(in %) is greater than or equal to 20%, 0 otherwise
age	Age of firm in years
gov_customer	Dummy=1 if principal buyer of output is government or government agencies, 0 otherwise
competitors_1	No Competitors
competitors_2	1 Competitor
competitors_3	2-5 Competitors
competitors_4	6+ Competitors
trade	Dummy=1 if firm imports or exports directly, 0 otherwise
electric	Dummy==1 if firm receives state provided electricity for production
water	Dummy==1 if firm receives state provided water supply for production
infraserv	Index (0-2) of availability of public services (electricity&water). Sum of "electric" and "water" variables.
security_naira	Annual cost of security in '000 Naira
regulation_realtime	Hours of senior mgt's time spent dealing with govt regs each week
external_consultant_naira	Amount('000Naira)spent on external consultants hired to deal with government regulations

equipment[22]. This is a measure of the degree to which capital is sunk. A positive value indicates that the capital stock is mobile. In order to make the interpretation of this residual easier, the value is multiplied by -1, so that a positive value represents a higher degree of sunk cost (capital is less reversible).

Also included as an explanatory variable is a dummy variable indicating whether the government is the principal buyer of the establishment's output. A bribe can be paid in order to secure a government contract that makes a company the sole supplier of a good that the government uses. Even in the absence of this situation, a bribe can be paid in order to get the public official to carry on choosing that particular company as a supplier.

A variable denoting the number of tax exemptions that the company is in receipt of is also included as a control rights measure. This variable is an index created by summing 4 dummy variables each denoting exemption from: import

22 $\frac{resale_i}{replace_i} = \gamma_0 + \gamma_1 * log(age) + \kappa_i$

duties; export taxes; profit tax; and Value Added Tax, respectively. A value of 1 indicates that the firm receives exemption from the respective tax scheme. In order to receive exemption from a tax, a company must approach the Federal Inland Revenue Service with its financial statements and incorporation documents. This encounter might also create the possibility for a bribe to be exchanged.

The NBS data includes a series of variables denoting operations performed by companies which required contact with public officials. These are listed in Table 3.4 and include: clearing goods through customs; obtaining road worthy certificates; procurement of goods and services from the government; obtaining business licenses and permits; getting clearance for environmental or sanitary regulations; obtaining residence and work permits; registering a vehicle; being involved with police investigations; having committed traffic offences; and being in contact with the courts. These variables serve as measures of the control rights hypothesis since they measure, to some extent, ways in which public officers might exercise control over the firms operations. The data also includes information on whether or not a bribe was paid for each of these activities. This allows for an analysis into the the propensity to bribe conditional on a specific meeting; for many different types of meetings with public officials. This builds on the previous data set by allowing for a heterogeneity of public officers and contact with them. Looking at different potential bribe attracting activities allows for a distinguishing of which activities

Table 3.2: Data Description (cont'd)

Variable Name	Definition
sales_t_1	Sales in previous year ('000 Naira)
sales_t_4	Sales 4 years previously ('000 Naira)
profit	Profit ('000 Naira) =sales - (operating costs+interest payments)
capital_stock	Resale value of machinery, vehicles & equipment ('000 Naira)
net_book_value	Net Book Value of capital stock (in '000 Naira)
replace_value	Cost of replacing capital stock with new equipment
un_sunk_cost	Residual from regression: $\frac{capital\_stock}{replace\_value} = log(age) + \epsilon$
alternative_return	capital_stock*sunk_cost
profit_employee	Willingness to pay: log(profits per ft employee)
ebribe_naira	Bribe per employee
eprofit	Profit per employee
ecapital_stock	Capital Stock Per Employee
eun_sunk_cost	Un_sunk_cost Per Employee
ealternative_return	Alternative Return Per Employee
esecurity_naira	Annual cost of security (in '000 NGN) per employee
eexternal_consultant_naira	Amount ('000 NGN) Spent On External Consultants Per Employee
external_auditor	Dummy=1 if annual financial statements checked by external auditor, 0 otherwise
university	Dummy=1 if owner/majority shareholder has a university degree, 0 otherwise
capital_labour	K/L ratio (labour=ft employees)
ind_loc_eprofit	Industry-location profit('000 Naira) per employee averages
electricity	Cost of electricity ('000 NGN)
pay_tax	log of (1+ percentage of sales reported for tax purpses)
african	Dummy=1 if sole owner/majority shareholder is african, 0 otherwise

Table 3.3: Data Description - NBS Data

Type	Variable Name	Definition	Measurement
	advantage_informal	advantage of informal over formal business	
Frequency And Knowledge Of Different Types Of Bribery	bribe_offer_official	A company offers gifts or money to a public official, directly or indirectly, in order that the official, in the exercise of his/her official duties, behaves in a way to facilitate the obtaining of what the company desires.	(Very Frequent; Fairly Frequent; Not Very Frequent But Not Unusual; Never Happens)
	bribe_demand_official	A public official asks a company for gifts, money or career advantages for his/her relatives, in order to depart from his/her normal behaviour in the exercise of his/her official duties	
	bribe_offer_foreign_official	A company offers gifts or money to a foreign public official, directly or indirectly, in order that the official, in the exercise of his/her duties, behaves in a way to facilitate the obtaining of what the company desires	
	bribe_demand_foreign_official	A foreign public official asks a company for gifts, money or career advantages for him/her or his/her relatives, in order to depart from his/her normal behaviour in the exercise of his/her official duties	

are most likely to attract bribes, and which are less so.

The NBS data also has information on the different ways in which bribery takes place. That is, whether a firm offers an informal payment to a public officer, or whether a public official demands a payment from a firm in order to carry out some special favour. This data allows one to decipher how bribery actually takes place: whether the firm offers or the official demands a bribe payment. With this information it should be easier to determine the direction of causality with regards to the payment of bribes; it should be clear whether the market is created by public official wanting more money; or by private firms wanting to bypass or speed up regulations[23]. The data also includes these questions for foreign public officials, which allows one to determine who is more susceptible to engage in corruption with Nigerian firms: foreign public officials or Nigerian ones[24].

Gauging The Level Of Bribery The wording on the questions concerning bribery and tax evasion in the Enterprise Survey are indirect. Rather than asking how much a firm pays in bribes the surveys queries how much similar firms pay. This method of asking indirectly is thought to generate more candor on the part of the interviewees about the bribes of their own firm. This is because the indirect nature of the questions deflects any guilt away from the individual/firm.

[23] It should be noted that there is the potential for reverse causality in this framework. I.e. a firm which has been previously demanded for a bribe by an official might tend to offer a bribe more often. Also, firms being quick to offer bribes might make public officials more likely to demand them from all firms.

[24] Comparisons of the behaviour of public officials from different countries is performed by Fisman & Miguel [2007] and Lambsdorff [1999, 2007].

Table 3.4: Data Description - NBS Data (cont'd)

Category	Variable Name	Definition
	bribe_dummy	Dummy=1 if firm admitted to bribing; 0 otherwise
	bribe_dummy_direct	Dummy=1 if firm paid a bribe in 2006; 0 otherwise
	bribe_dummy_indirect	Dummy=1 if firm indirectly admitted to bribing; 0 otherwise
	bribe_naira	amount paid as bribe ('000 Naira)
	zone	Geo-Political Zone (NC;NE;NW;SE;SS;SW)
	foreign	Foreign ownership =25%
	security	Dummy=1 if company using security service
	age	Age of firm (years)
	lage	Log of age
	trade	Dummy=1 if firm engages in international trade, 0 otherwise
operation performed by companies/ bribe paid given that the company performed the operation	conda/bribea	clearing goods through customs
	condb/bribeb	obtaining road worthy certificates
	condc/bribec	procurement of goods and services from government
	condd/bribed	obtainig business licenses and permits
	conde/bribee	procurement of goods and services from private companies
	condf/bribef	getting clearance for environmental or sanitary regulations
	condg/bribeg	residence and work permits
	condh/bribeh	vehicle registrations
	condi/bribei	police investigations
	condj/bribej	traffic offences
	condk/bribek	contact with the court

This method of asking indirectly is thought to be useful in gauging the level of bribe paid by the firm due to the false consensus effect [Ross, Greene & House , 1977]. This is the phenomenon that people who engage in a socially undesirable act are more likely to overestimate the prevalence of that act amongst their peers. In this case, firm managers who pay bribes are more likely to say that similar firms also pay bribes. This helps the current paper to interpret the answer to indirect questions as being representative of the firm itself.

The indirect form of posing questions is similar to that used on the 1998 Enterprise Survey for Uganda [Reinikka & Svensson , 2003]; the Business Environment And Enterprise Performance Survey (BEEPS) on the transition countries [Hellman, Jones & Kaufmann , 2000]; and the North-African Survey of Firms conducted by Université Paris 1 [Delavallade , 2011]. This allows for comparability between different countries.

Further support for the validity of the data concerning informal payments is provided by the results of similar surveys carried out within Nigeria and in other countries. In a survey of businesses in Romania [Azfar & Murrell , 2009], there is no difference in the admitting of bribery between firms asked about their own businesses and those asked about "businesses like yours". This lends favour to the argument that no less information is retrieved by asking bribery related questions indirectly.

3.4 Empirical Specification Of The Control Rights And Bargaining Hypotheses

3.4.1 Empirical Models To Test The Control Rights And Bargaining Hypotheses

The control rights theory can be formally described as:

$$d_i = \eta' C_i + u_i \tag{3.4.1.1}$$

where d_i is the probability that company i pays a bribe; C_i is a vector of variables denoting mandatory interactions with public sector officials; η' is a vector of coefficients; and u_i is an error term.

The bargaining theory can be stated as:

$$b_i = \alpha_0 + \alpha_1 \pi_i(k) + \alpha_2 E\pi_i(k) + \alpha_3 E\pi_i(\kappa k) + \epsilon_i \tag{3.4.1.2}$$

where b_i is the amount of bribe paid by company i; $\pi_i(k)$ is current profit; $E\pi_i(k)$ is expected profit in the next period; $\pi_i(\kappa k)$ is expected return if the company reallocates its capital in another market; and ϵ_i is an error term. Based on the preceding analysis (section 3.2), α_1 and α_2 are expected to be positive while α_3 is expected to be negative.

In order to derive an equation describing the bribe behaviour of firms this analysis formulates this problem into that of a latent variable model, describing the propensity to bribe, consisting of a selection equation and an outcome equation. The selection equation describes whether or not the firm pays a bribe and the outcome equation describes the amount of bribe paid by firms that pay a bribe.

Selection:

$$y_1 = \begin{cases} 1 & if y_1^* > 0 \\ 0 & if y_1^* \leq 0 \end{cases} \tag{3.4.1.3}$$

Outcome:

$$y_2 = \begin{cases} y_2^* & if y_1^* > 0 \\ - & if y_1^* \leq 0 \end{cases} \tag{3.4.1.4}$$

where y_1^* is the (unobserved) propensity to pay a bribe; y_1 is a dummy variable equal to 1 if the firm pays a bribe, 0 otherwise; and y_2 is the observed bribe payment.

This lends itself to the use of a selection model:

$$p[bribe] = \mathbf{x_1'} \beta_\mathbf{1} + \epsilon_1 \tag{3.4.1.5}$$

$$y_2 = \mathbf{x_2'} \beta_\mathbf{2} + \epsilon_2 \tag{3.4.1.6}$$

Where ϵ_1 and ϵ_2 are possibly correlated. This paper assumes that ϵ_1 and ϵ_2 are joint normally distributed and homoskedastic. That is:

$$\epsilon_1 \sim N(0, \sigma_1^2) \tag{3.4.1.7}$$

$$\epsilon_2 \sim N(0, \sigma_2^2) \tag{3.4.1.8}$$

and $Cov(\epsilon_1 \epsilon_2) = \sigma_{12}$. The value of σ_1^2 is normalised to 1 because the value for y_1^* is not observed but the sign is. The likelihood function for this model is:

$$L = \prod_{i=1}^{n} [Pr(y_{1i}^* \leq 0)]^{1-y_{1i}} [f(y_{2i}|y_{1i}^* > 0) \times Pr(y_{1i}^* > 0)]^{y_{1i}} \tag{3.4.1.9}$$

where the first term represents the contribution to the likelihood function when the propensity to bribe is less than or equal to zero; and the second term represents the contribution when the propensity to bribe is positive. This model is estimated using a censored Tobit model with density function:

$$f(y_{2i}) = \left[\frac{1}{\sqrt{2\pi\sigma^2}} exp \left\{ -\frac{1}{2\sigma^2} (y_{2i} - \mathbf{x'_{2i}} \beta^2) \right\} \right]^{b_i} [\Phi \{ (-\mathbf{x'_{2i}} \beta)/\sigma_2 \}]^{1-b_i} \tag{3.4.1.10}$$

Where $b = 1$ if the firm pays a bribe, 0 otherwise.

From Equation 3.4.1.5; $E(y_2|\mathbf{x}, y_1^* > 0) = \mathbf{x'_2}\beta_2 + \mathbf{E}(\epsilon|\mathbf{y_1^*} > \mathbf{0})$. Given the assumption that the errors are normally distributed, then: $E(\epsilon|y_1^* > 0) = \sigma_{12}\lambda(\mathbf{x'_1}\beta_1)$

The expected value of bribe to be paid, y_2, for the non-censored observations is therefore given by:

$$E(y_2|\mathbf{x}, y_1^* > 0) = \mathbf{x'_2}\beta_2 + \sigma_{12}\lambda(\mathbf{x'_1}\beta_1) \tag{3.4.1.11}$$

where $\lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)$, ϕ is the standard normal density function; and Φ is the standard normal cumulative distribution function. An OLS regression of y_2 on x_2 would suffer from omitted variable bias due to the missing variable ϕ_i/Φ_i, this would yield an inconsistent estimate of β_2. This missing variable can be generated by collecting the pseudo-residuals from a probit model of y_1 on $\mathbf{x'_1}$. An OLS regression of y_2 on $\mathbf{x_2}$ and $\lambda(\mathbf{x'_1}\beta_1)$ provides a semiparametric estimate of β_2 and σ_{12}. If the nonselection hazard $\lambda(\mathbf{x'_1}\beta_1)$ is revealed to be insignificant, then an OLS regression on the uncensored sample (bribing firms) will yield consistent estimates of β_2.

In describing the bribe behaviour of firms, it may be the case that the factors which determine bribery also determine the amount of bribe paid. In such a case, all companies who do not pay a bribe are thought to pay a bribe of amount zero.

Due to the censored nature of the data on bribes (the amount of bribe paid cannot be negative), such analysis can be performed using a censored Tobit model, this would allow for the inclusion of all firms in the bribery equation. A characteristic of this model is that it restricts the sign of the effect of a covariate on the probability of paying a bribe to be the same as the effect of the variable on the amount of bribe paid.

Officials in a heavily regulated sector-region might demand more bribes from firms, however, because of this, the amount of bribe that they require from each company might be less than other industry-locations. In such a case the sign for a dummy that shows whether or not a firm is located in that industry-location would be positive for the selection equation (propensity to bribe) and negative for the outcome equation (amount of bribe paid). This restriction is tested using a likelihood ratio-test (Tobit specification test), which is described in section 3.9.

In the absence of selection bias (if the non-selection hazard term in 3.4.1.11 is insignificant) then a two-part model can be used to estimate the factors influencing the occurence of bribery and the amount of bribe that is paid. This model is composed of a probit estimation on all firms, and an ordinary least squares regression on those firms that report a positive bribe payment. In the absence of selectivity the two-part model is more robust than the full-information maximum likelihood or two-step Heckman procedure [Puhani , 2000].

In the two-part model the propensity to bribe is estimated as a probit model:

$$Pr[BribeDummy_i = 1] = \Phi(\eta_C' C_i + \alpha_B' B_i) \tag{3.4.1.12}$$

where $[BribeDummy_i = 1]$ is the case where a firm reports a positive amount of bribe. C_i are the variables in line with the control rights hypothesis: the receiving of public services; whether or not the firm imports or exports directly; and the share of sales that the firm reports for tax purposes. B_i represents the variables in line with the bargaining hypothesis: current profit; expected future profit; and the alternative return on the firms capital stock. Φ is the standard normal distribution function.

The second part of the two-part model estimates the amount of bribe that is paid using a reduced sample ordinary least squares regression:

$$Bribe_i = \beta_0 + \beta_1 Profit_i + \beta_2 E[Profit_{i,t+1}] + \beta_3 AlternativeReturn_i + \beta_c' \mathbf{C_i} + v_i \tag{3.4.1.13}$$

where $Profit_i$ denotes the current profit of the firm; $E[Profit_{i,t+1}]$ represents expected future profit; and $AlternativeReturn_i$ is the expected return if the firm reallocates its capital for use in another sector. $\mathbf{C}$ is the vector of variables representing the control rights hypothesis.

Empirical Models To Analyse The Propensity To Meet With A Public Official And The Conditional Probability Of Bribing A Public Official
The NBS data used in this study has information concerning whether or not a company met with a public official in the previous year; and whether or not a firm paid a bribe to a public official during the previous year. Coming into contact with a public official is necessary in order to bribe that public official, therefore, not coming into contact with a public official is a perfect predictor of not bribing a public official. The discussion in subsection 3.2 suggests that whether or not a company comes into contact with a public official depends upon the industry that the firm is located in, whilst the payment of a bribe to an official depends on firm specific factors.

In order to investigate the factors influencing the coming into contact with public officials and the payment of bribes to public officials one might use two separate binary outcome models. However, this would impose the assumption that the unobservable factors that influence the propensity to meet with a public official are independent of the unobservables that influence the propensity to bribe a public official [Dolton, 2012][25]. For example, a new piece of legislation that imposes a fee on firms that produce over a certain level of CO_2 emissions might lead firms to pay the official in charge a bribe (that is less than the fee) in order to bypass the paying the fee. Estimating two separate binary outcome models would not allow for this correlation between the two processes of meeting with a public official and of bribing a public official

In order to relax the assumption that the errors from the two equations are independent of each other this study uses a bi-variate probit analysis to analyse the meeting with and paying of bribes to public officials. However, in order to fit the current data, the model must allow for the fact one does not observe the bribe payments of the companies that do not come into contact with public officials. Therefore, this chapter uses a bivariate probit model with sample selection[26]. This model allows for both selection on observed regressors and unobserved errors. In the current study this model is formed of a participation equation which models the propensity (or decision) to meet with a public official:

$$y^*_{1j} = \mathbf{X}_{1ij}\beta 1j + u_{1ij} \tag{3.4.1.14}$$

where

$$y_{1j} = \begin{cases} 1 & if \quad y^*_{1j} > 0 \\ 0 & if \quad y^*_{1j} \leq 0 \end{cases} \tag{3.4.1.15}$$

and an outcome equation that models the propensity (decision) to pay a bribe

[25] Correspondence with Author

[26] This model has also been referred to as: double probit; censored probit; and bivariate probit with partial partial observability.

to a public official:

$$y^*_{2j} = \mathbf{X}_{2ij}\beta 2j + u_{2ij} \tag{3.4.1.16}$$

where

$$y_{2j} = \begin{cases} y^*_{2j} & if \quad y^*_{1j} > 0 \\ - & if \quad y^*_{1j} \leq 0 \end{cases} \tag{3.4.1.17}$$

and $Cov\left[u_{1ij}, u_{2ij}\right] = \rho \neq 0$, or:

$$\begin{aligned} u_{1ij} &= \eta_{ij} + \epsilon_{1ij} \\ u_{2ij} &= \eta_{ij} + \epsilon_{2ij} \end{aligned} \tag{3.4.1.18}$$

Where y_{1j} is a dummy variable equal to 1 if the firm met with public official type j, 0 otherwise; and y_{2j} is a dummy variable equal to 1 if the firm paid a bribe to public official type j, 0 otherwise. In this case, the probability of bribing official type j, is equal to the probability of meeting with offical type j times the probability of bribing official type j conditional on meeting with official type j. Formally:

$$P(y_{2j} = 1) = P(y_{1j} = 1) * P(y_{2j} = 1 | P(y_{1j} = 1)) \tag{3.4.1.19}$$

or

$$P(y_{2j} = 1) = P\left[y^*_{1ij} > 0\right] * P\left[y_{2ij} = 1 | y^*_{1ij} > 0\right] \tag{3.4.1.20}$$

There are three types of observations within this framework: those who do not meet with public officials and therefore do not pay any bribes ($y_{1j} = 0$); those who meet with public officials and do not pay any bribes ($y_{1j} = 1$ & $y_{2j} = 0$); and those who meet with public officials and pay bribes ($y_{1j} = 1$ & $y_{2j} = 1$). The respective probabilities of each of these events can be described as: $P\left[y_{1j} = 0\right] = \Phi(-\mathbf{x_1'}\beta_1)$; $P\left[y_{1j} = 1, y_{2j} = 0\right] = \Phi(\mathbf{x_1'}\beta_1) - \Phi_2(\mathbf{x_1'}\beta_1, \mathbf{x_2'}\beta_2, \rho)$; and $P\left[y_{1j} = 1, y_{2j} = 1\right] = \Phi_2(\mathbf{x_1'}\beta_1, \mathbf{x_2'}\beta_2, \rho)$.

The log likelihood for this model is:

$$\begin{aligned} lnL = & \sum_{i=1}^{N}(1 - y_{1ij}) ln\Phi(-\mathbf{x_1'}\beta_1) \\ & + y_{1ij}(1 - y_{2ij}) ln\left[\Phi(\mathbf{x_1'}\beta_1) - \Phi_2(\mathbf{x_1'}\beta_1, \mathbf{x_2'}\beta_2, \rho)\right] \\ & + y_{1ij} y_{2ij} ln\Phi_2(\mathbf{x_1'}\beta_1, \mathbf{x_2'}\beta_2, \rho) \end{aligned} \tag{3.4.1.21}$$

3.5 Comparing The Two Datasets

This section compares the respective companies in the enterprise and NBS surveys. It looks at the proportion of bribe paying firms and the average amount of bribes paid for each sample and performs a test to see if there is a significant difference between the two datasets.

Results - Descriptive Statistics Firms in the Enterprise Survey showed a greater willingness to report bribery than those in the NBS survey. 52% of firms in the ES reported a positive bribe payment compared to 27% of the NBS sample (Table 3.5). Nothing changes when focusing on the manufacturing firms within the respective datasets.

Table 3.5: Prevalence Of Bribery For Both Datasets

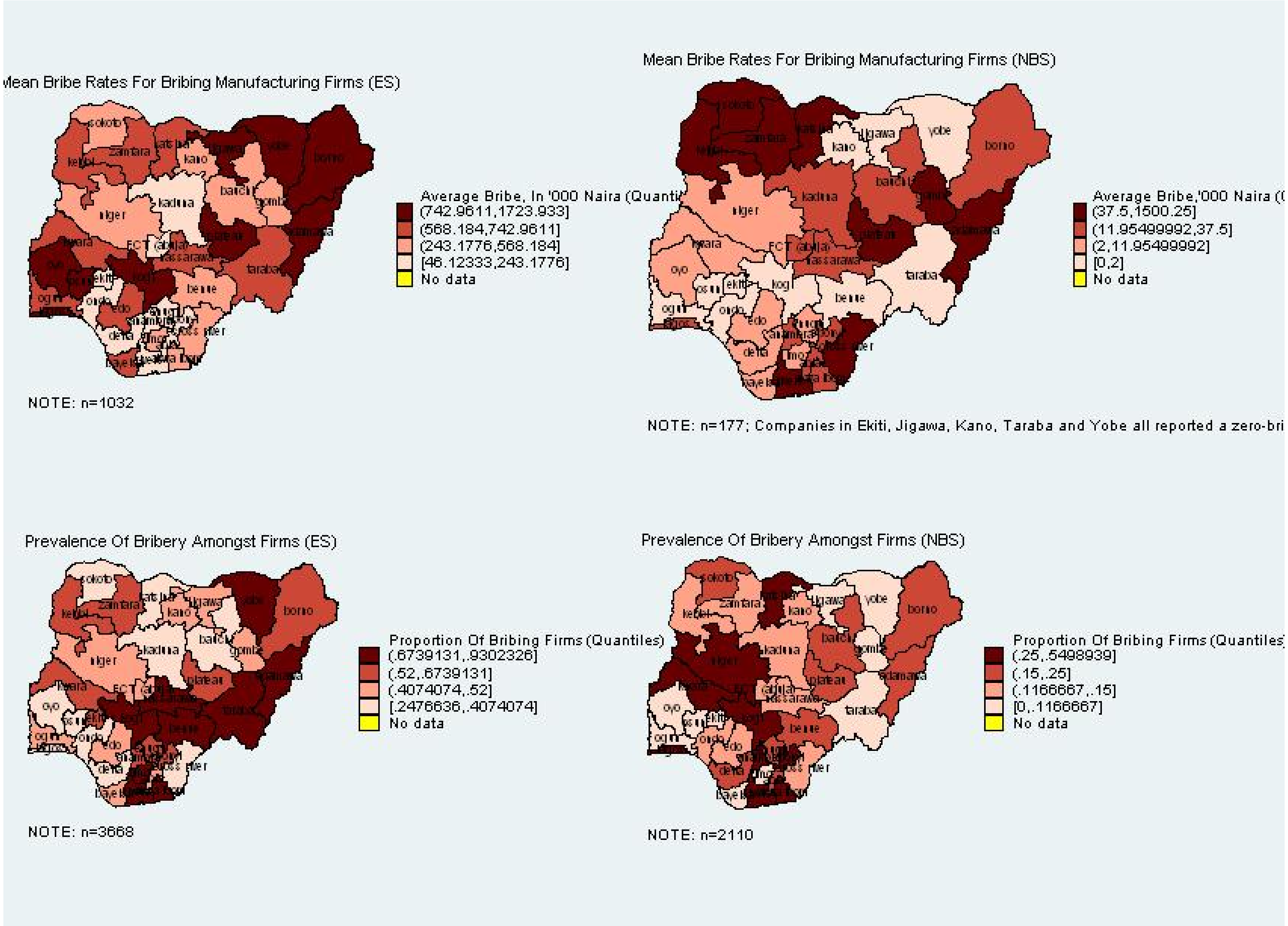
Dataset	Obs	Mean	Std. Dev.	Min	Max
WB	3668	.52	.50	0	1
NBS	2110	.27	.44	0	1

The average bribe payment for the ES full sample (All Industries) was ₦293,400. Focusing on bribe paying firms alone raises this average to ₦562,200 (Table 3.6). Table 3.6 also shows a wide disparity in the average reported bribe payments between companies who report in terms of the absolute Naira amount and those who choose to report their bribe payments as a percentage of sales. Focusing on firms who reported paying a bribe (the lower subsection of table 3.6), the average reported bribe payments amongst sales reporters was ₦690,000; whereas those who reported their bribe as a Naira figure reported, on average, ₦73,700. The distribution of bribe payments for firms in the NBS sample appears to be closer to that of the ES Naira reporters: the average reported bribe payment for this sample of firms is ₦53,400. These results are also seen when looking solely at the manufacturing industry in Nigeria. Average bribe payments for bribing sales reporters, Naira reporters, and NBS firms , respectively, are: ₦710,300; ₦85,500; and ₦77,900.

Table 3.6: Distribution Of Bribe Payments (In ’000 Naira) For Both Datasets

Dataset	Obs	Mean	Std. Dev.	Min	Max
WB (All Firms)	3668	293.4	667.4	0	5880
WB (Bribing Firms)	1914	562.2	838.2	1	5880
WB(Sales Reporters - All Firms)	3246	322.49	703.5	0	5880
WB (Sales Reporters - Bribing Firms)	1517	690	897.5	1.2	5880
WB(Naira Reporters - All Firms)	422	69.353	92.29586	0	768
WB (Naira Reporters - Bribing Firms)	397	73.7	93.5	1	768
NBS (All Firms)	2110	4.5	73.9	0	3000
NBS (Bribing Firms)	177	53.4	250.5	.01	3000

Figure 3: Comparison Of The Prevalence Of Bribery; And The Average Bribe Rates Amongst Bribing Manufacturing Firms (in '000 Naira) For Both Datasets (NBS & WB)



The regional distribution of bribe incidence and the average bribe payments per state are shown in figure 3. The lower part of the diagram corresponds to table 3.5. It shows the proportion of bribe reporting firms across different states in Nigeria for all firms (all industries) of the ES and NBS samples. Results indicate a relatively higher prevalence of bribing firms in the South-South region of Nigeria. These include Rivers, Imo and Akwa Ibom and these results are consistent between both datasets. Kogi State also has a relatively high prevalence of bribery. Focusing on the average bribe payments for bribe reporting manufacturers, companies in Adamawa appear to consistently report a higher bribe payment in both samples.

Perceptions Versus Reality Perception based indicators of corruption can suffer from bias [Carlin & Seabright , 2007]. More productive companies have a higher valuation of the business environment than productive firms. Accordingly, any constraint to the business environment serves as a higher cost to operations compared with less productive companies. Following from this, relatively more productive firms are more likely to complain about constraints to the business environment than less productive ones. Therefore creating a bias in perception indices of the state of the business environment.

One possible solution to this, in the case of bribery, is to compare subjective reports on the extent of corruption to the actual reported bribe payments. If all firms face the same business environment, and in the absence of misreporting and bias; companies in areas where bribery is more pervasive should report this in their subjective valuations of the business environment.

Companies in the Enterprise Survey were asked: "Do you think that the following present any obstacle to the current operations of your establishment?" with "corruption" being one of the options.[27]

Results from table 3.7 shows some evidence that businesses are revealing something when they give subjective evaluations of the business environment. Only 39% of those who said that corruption presented "no obstacle" to business operations paid a bribe to deal with business regulations whereas 69% of those who reported that corruption posed a "very severe obstacle" to operations reported paying a bribe. The values for the proportion of bribing firms increases as one moves across the table from 'no obstacle' to 'very severe obstacle'. The proportion of bribing firms decreases as one moves from the 'moderate obstacle' group to the 'major obstacle' group, however, this difference is not statistically significant at the 10% level. This pattern is also present in the NBS dataset.

[27] To be sure, the term "corruption" encompasses a number of acts, so companies might have been referring to bribery amongst other things when responding to this question. However, the data in table 3.7 provides some information as to the relationship between bribery and the other elements that come under the term "corruption".

Table 3.7: Proportion Of Bribing Firms And Subjective Reports On Corruption As An Obstacle To Operations-ES

Bribing Behaviour	Corruption As An Obstacle to Business Operations					Total
	No Obstacle	Minor Obstacle	Moderate Obstacle	Major Obstacle	Very Severe Obstacle	
Non-Bribers (Row %)	35.60	20.78	20.06	16.67	6.89	100.00
Non-Bribers (Column %)	60.81	52.06	41.58	44.88	31.31	48.58
Bribers (Row %)	21.67	18.08	26.63	19.34	14.29	100.00
Bribers (Column %)	39.19	47.94	58.42	55.12	68.69	51.42
Total (Row %)	28.44	19.39	23.44	18.04	10.69	100.00
Total (Column %)	100	100	100	100	100	100

This non-decreasing relationship between observed corruption and reported bribery is also present in the NBS dataset.

Different Types Of Bribe Results from sections 3.6 to 3.10 describe a positive correlation between firms meeting with officials and firms paying bribes to officials. The story that seems to be coming out of the data is that having to meet with a public official is related to a higher probability of paying a bribe (to a public official) to speed regulation or assist with business operations in some other way.

A potential limitation of the previous analysis is that it does not identify the exact purpose for which the bribe was paid. A firm might regularly meet with public officials from different arms of the government (e.g. police; customs; the courts) but might only pay bribes to some of them and not others. The preceding analysis does not allow one to identify which public officials have been bribed out of all of the officials that have been met with. Tables A.1 to A.11 (in Appendix Section A.1) present the results of the Bivariate Probit Analysis of the bribery of public officials that distinguishes between different types of public officials.

3.6 Summary Statistics From The Enterprise Survey

This section discusses the results from the data. It provides an analysis of the factors affecting whether or not a company pays a bribe, and the amount that the firm has to pay. Summary statistics for the Enterprise dataset are shown in Tables 3.8 and 3.9. 80% of firms were small in size (5-19 full-time permanent employees); 19% were medium sized (20-99 employees); and 1% of firms had more than 100 employees. The mean age of the firms is 14 years. Less than 1% of firms were foreign owned but roughly 2% of the sample engaged in some form of international trade (either exporting directly or importing directly)[28]. 25% of firms received government provided water in their production process, the figure for electricity was 2%. On average, the firms operate for 62 hours per week, (on average) 3 of which are spent by senior managers dealing with government regulations.

[28] 1.1% of firms solely imported; 0.45% of firms solely exported; and 0.15% of firms engaged in both importing and exporting

Table 3.8: Correlation Matrix

	(1)															
	bribe_naira	bribe_dummy	profit	employee_total	gov_customer	trade	infraserv	african	foreign	regulation_realtime	tax_form_time	tax_visit	capital_stock	un_sunk_cost	tax_percentage	electricity
bribe_naira	1															
bribe_dummy	0.4***	1														
profit	0.3***	0.0003	1													
employee_total	0.2***	-0.04	0.3***	1												
gov_customer	0.06*	0.04	0.009	0.0003	1											
trade	-0.0009	0.004	0.05*	0.07**	-0.02	1										
infraserv	0.08***	0.07**	0.1***	0.06**	-0.02	0.01	1									
african	-0.04	-0.06**	0.005	-0.03	0.003	-0.05*	0.02	1								
foreign	-0.02	-0.03	0.05*	0.02	-0.006	0.07**	0.05*	-0.04	1							
regulation_realtime	0.04	0.05*	0.04	0.09***	0.02	0.006	0.07***	-0.02	0.008	1						
tax_form_time	-0.06**	-0.09***	0.05*	0.1***	-0.01	0.02	-0.04*	0.02	0.008	0.05*	1					
tax_visit	-0.01	-0.08***	0.04	0.07**	-0.04	0.03	-0.02	-0.003	-0.005	-0.03	0.02	1				
capital_stock	0.1***	0.04	0.2***	0.2***	0.03	0.09***	0.05*	-0.008	0.01	0.03	-0.02	0.03	1			
un_sunk_cost	-0.10***	-0.09***	0.005	-0.02	-0.05*	0.02	0.08***	0.03	-0.006	0.06**	0.06**	0.03	-0.05*	1		
tax_percentage	0.06**	0.1***	0.03	-0.007	-0.02	0.008	0.02	0.04	-0.04	0.02	-0.07**	0.1***	0.02	-0.02	1	
electricity	0.2***	0.1***	0.1***	0.2***	0.01	-0.007	0.08***	0.02	0.04*	0.01	-0.004	0.03	0.2***	-0.2***	0.1***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.9: Summary Statistics, by bribe reporting behaviour

Variable	Statistic	All Firms	Bribe=0	$Bribe > 0$	$Bribe > 0$ (Sales %)	$Bribe > 0$(Naira)
bribe_naira	Mean	294	0	572	710	86
	Median	8	0	264	390	50
	Std. Dev.	652	0	816	876	87
ebribe_naira		20.6	0.0	40.0	49.4	6.6
		0.5	0.0	22.2	32.3	3.8
		49.6	0.0	63.3	68.7	7.5
employee_total		17	18	16	16	17
		12	13	12	12	13
		16	18	14	13	14
eprofit		158.7	155.3	161.8	167.2	142.6
		90.0	77.8	100.1	102.0	95.8
		216.8	211.8	221.4	233.9	169.3
capital_labour		230.2	190.9	267.2	282.1	214.5
		100.0	75.0	125.0	131.4	100.0
		718.8	544.4	850.0	943.0	362.3
un_sunk_cost		-0.88	-0.88	-0.89	-0.88	-0.89
		-0.88	-0.88	-0.89	-0.88	-0.90
		0.03	0.02	0.03	0.03	0.03
infraserv		0.29	0.25	0.32	0.32	0.32
		0.00	0.00	0.00	0.00	0.00
		0.50	0.48	0.51	0.52	0.47
tax_percentage		69.6	66.1	72.8	70.5	80.9
		75.0	70.0	75.0	75.0	100.0
		27.4	28.2	26.2	25.7	26.3
trade		0.02	0.02	0.02	0.01	0.04
		0.00	0.00	0.00	0.00	0.00
		0.13	0.13	0.13	0.11	0.18
N		2001	972	1029	802	227

In each column, for each variable, the table reports the mean, median and standard deviation for each (sub)sample. The last row details the number of observations for each (sub)sample.

The average annual cost of dealing with requirements imposed by the government was ₦464,000, of which ₦48,500 was spent on external consultants used to deal with regulations. 83% of firms were visited, inspected by, or required to meet tax officials in the previous year, and firms took an average of 16 hours to fill in all forms and requirements to pay local taxes. The average amount of sales that firms reported for tax purposes was 70%, compared with 61% of employees reported for the same purpose. The average level of total employment (full-time; part-time; and temporary/seasonal) was 17 workers. 9% of firms had their financial statements checked by an external auditor and 3% of business owners had a university degree.

Average sales declined from 4 years previously to the year before the survey was taken, from ₦11.3 million to ₦8.4 million. The average value of machinery, vehicles and equipment for the manufacturing industry was ₦3.4 million. The average annual cost of electricity was ₦148,000.

Bribe Reporting The correlation matrix in Table 3.8 shows a statistically significant positive relationship between the receipt of public services ("infraserv") and the payment of bribes ("bribe_dummy"), with a correlation of 0.07 which is significant at the 5% level. The percentage of sales reported for tax purposes ("tax_percentage") is also significantly related to the payment of bribes, with a correlation of 0.1 that is significant at the 1% level. Both variables are also significantly positively correlated with the amount of bribe paid (0.08 [1%] and 0.06 [5%], respectively). Profits and size of firm (employees) are also positively correlated with the bribe payment (0.3 and 0.2, respectively, both significant at the 1% level).

Table 3.9 shows summary statistics of the main control rights and bargaining variables for all firms, and then disaggregated by the reporting behaviour of the firm: firms either reported no bribe; a positive bribe as a percentage of total sales; or a positive bribe in Naira. There were no non-respondents to the bribery question in this sample. Firms were allowed to report their level of graft as a percentage of sales or in Naira. The majority of firms who reported a positive bribe chose to report in terms of sales percentage (802 versus 227). Mean comparison (Student's t) tests show that firms reporting in terms of sales report significantly higher values of bribe payments, bribe payment per employee, profit and amount of moveable capital than Naira reporters. While Naira reporters declare a larger proportion of their sales for tax purposes and (on average) engage in more international trade than sales reporters[29].There are no significant differerences, however, in their number of employees; capital per worker; or their receipt of public services. The median bribe for the firms reporting bribes in Naira amounts were ₦50,000 whilst this value for sales reporters was ₦390,000.

Results show that 51% of companies admitted to paying a bribe (Table 3.9). The sample contained no non-respondents to the bribery related question. Amongst companies that reported a positive bribe, the average amount paid was ₦572,000[30] per year with a median of ₦264,000 (£1050). This corresponds to ₦40,000 per employee (or 6% of operating costs). Looking at the entire sample reduces these averages to ₦294,000; ₦20,000 per employee; and 3.3% of operating costs, respectively. Out of the firms that reported a positive bribe amount; 22% reported this in Naira and 78% reported this as a percentage of sales.

Chemical manufacturing and wood manufacturing had the highest within-industry percentage of bribe paying firms[31] (57% and 56% respectively). Electronics and garment manufacturing were the least graft-intensive industry subsectors

[29]Results for bribe; bribe per employee; moveable capital; and tax percentage are significant at the 1% level of significance. The result for international trade is significant at the 5% level.

[30]£2270 in October 2007(£1:₦251.97)

[31]Excluding "Other Manufacturing" with a high of 67% firms reporting graft.

with 17% and 43% of firms reporting bribes in each industry respectively[32]. The most bribe-intensive states were Benue; Taraba; and Yobe, with each state having all of its firms admitting to paying some bribe. Lagos and Cross River state had the lowest percentage of bribing firms (19% and 27% respectively).

Sample Selection A potential concern with the analysis is sample selection bias. Equations 3.4.1.12 and 3.4.1.13 can be interpreted as a selection model. If the error terms in these two equations are correlated, an OLS regression on equation 3.4.1.13 will return biased estimates. The selection model can be estimated by a two-step procedure where 3.4.1.12 is estimated by probit and 3.4.1.13 with OLS; and using an estimate of the inverse Mills ratio from the first step to correct for selection bias. This study chooses to identify the model by excluding public services and trade from the second stage of the estimation. The support for this exclusion restriction is based on section 3.2, that interactions with public officials do not affect the amount of bribe paid. Table 3.10 presents the results from the two-step procedure and compares it with the OLS estimates. The top half of the table displays results from the second stage of the two-step procedure while the bottom half has results from the first stage.

In the first stage analysis the variables for: having the government as the principal buyer of goods; and receiving public services have the largest positive coefficients on the probability of paying a bribe. The proportion of moveable capital and the number of competitors have the largest negative coefficients. Profit and capital per worker enter into the second stage model with significant coefficients while the percentage of sales reported for tax purposes and access to public services enter insignificantly. In all specifications the coefficient on the inverse-mills ratio enters insignificantly which suggests an absence of selection bias in the data. This also explains the similarity in bargaining coefficients between the Heckman model and OLS. Due to the absence of evidence for selection bias, the two-part model and Tobit can be used to analyse the factors determining the probability of paying a bribe and the amount of bribe paid.

The Two-Part Model The next sections elucidate the two-part model consisting of a probit model on the entire sample and a subsample OLS regression on the bribing firms alone. Due to the low correlation between the errors in the selection equation and outcome equation (and therefore absence of selection bias; [section 3.6]) and moderate degree of censoring, this model is more robust than the Heckman model.

[32] Only 0.3% of the sample (6 firms) were from the electronics manufacturing sector.

Table 3.10: Estimations Using Heckman Two-Step Procedure

	m1	m2	m3	m9 (OLS)
Outcome Equation				
eprofit	0.170***	0.170***	0.153***	0.094***
	(0.008)	(0.008)	(0.040)	(0.025)
capital_labour	0.006***	0.007***	-0.004	0.006**
	(0.002)	(0.002)	(0.017)	(0.002)
un_sunk_cost	-165.540*	-166.170*	552.696	
	(91.874)	(92.014)	(1076.681)	
competitors_4	-11.208	-11.872	55.515	-7.244*
	(10.272)	(10.300)	(103.909)	(4.210)
eexternal_consultant_naira	0.605***	0.597***	1.273	0.027
	(0.153)	(0.153)	(1.122)	(0.184)
african	-3.637	-4.017	54.564	-6.632**
	(9.132)	(9.151)	(90.450)	(3.041)
trade		-24.885**		-16.972***
		(11.805)		(6.381)
eun_sunk_cost				-88.986***
				(24.672)
Constant	-132.563	-134.154	689.522	5.397
	(97.258)	(97.407)	(1218.583)	(6.422)
Other control variables	YES	YES	YES	YES
Selection Equation				
infraserv	0.174***	0.174***	0.174***	
	(0.058)	(0.058)	(0.058)	
tax_percentage	0.006***	0.006***	0.006***	
	(0.001)	(0.001)	(0.001)	
capital_labour	0.000*	0.000*	0.000*	
	(0.000)	(0.000)	(0.000)	
un_sunk_cost	-4.260***	-4.260***	-4.260***	
	(1.127)	(1.127)	(1.127)	
competitors_4	-0.428**	-0.428**	-0.428**	
	(0.167)	(0.167)	(0.167)	
gov_customer	0.504*	0.504*	0.504*	
	(0.265)	(0.265)	(0.265)	
regulation_realtime	0.017**	0.017**	0.017**	
	(0.007)	(0.007)	(0.007)	
eexternal_consultant_naira	-0.004*	-0.004*	-0.004*	
	(0.002)	(0.002)	(0.002)	
african	-0.382***	-0.382***	-0.382***	
	(0.146)	(0.146)	(0.146)	
Constant	-3.530***	-3.530***	-3.530***	
	(1.027)	(1.027)	(1.027)	
Other control variables	YES	YES	YES	N/A
Inverse-Mills Ratio	16	17	-253	
	(28.1)	(28.2)	(391.4)	
Observations	2001	2001	2001	2001
Censored Observations	972	972	972	
Chi-Squared	711.3	713.4	38.2	
F-Stat				4.83

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dependent variable of selection equation is a dummy variable equal to 1 if the firm reported a positive bribe payment; 0 otherwise. Dependent variable of outcome equation is bribe amount in '000 Naira. Other control variables for the outcome equation include: percentage of sales reported for tax; whether or not the government is a customer (dummy); number of managers hours spent dealing with government regulations per week; and infraserv (category). The coefficients on these variables are all insignificant. Other control variables for the selection equation include: trade (dummy); and profit per employee. The coefficients on these variables are all insignificant.

3.7 Factors Determining Whether Or Not A Bribe Is Paid: First Stage Probit Analysis

The first set of results use the enterprise survey dataset while the second set of results use the NBS dataset. The main results from the two-part model are shown in tables 3.11 and 3.12. Results from robustness checks, using the Tobit model, are displayed in Table 3.13. Further analysis using the multinomial logit model is shown in Table 3.19. The main results using the NBS dataset are found in Tables 3.14 and 3.15. The former table deals with the factors affecting the propensity to bribe while the latter looks at the instigator of the bribe transaction.

Probit estimations representing equation 3.4.1.12 are shown in Table 3.11. All variables have their expected signs[33]. In addition, all variables retain their sign when entered sequentially in the order shown. Supporting the conceptual framework, firms that receive public services (*infraserv*), engage in international trade (*trade*), and report a higher percentage of their sales for tax purposes (*tax_percentage*) have a higher probability of having to give an informal payment or gift with regard to customs, taxes, licenses, regulations and services. The coefficients on public services and tax are statistically significant at the 1% significance level in every model. It thus appears that having to deal with government officials means that a firm is more likely to be under bureaucratic control and therefore more likely to have to pay a bribe.

The control rights measures represent contact with government officials (or a signal of being in the formal sector), despite this, these three variables are not highly correlated (Table 3.8) and can be used within the same model. The sign, statistical significance and magnitude of the coefficients on these variables do not change when they are jointly entered into a model compared to when they are entered using 3 seperate partial-effect models.

[33] "trade" fluctuates between positive and negative but remains insignificant for all specifications. Out of a total of 34 firms (1.7% of the sample) that engaged in the direct importing/exporting of goods, just over half (18) reported a positive bribe payment

Table 3.11: Probit Estimations On The Propensity To Bribe

	m1	m2	m3	m4	m12
infraserv	0.175***			0.170***	0.172***
	(0.057)			(0.057)	(0.067)
trade		0.039		0.023	-0.000
		(0.217		(0.216)	(0.241)
tax_percentage			0.006***	0.006***	0.004***
			(0.001)	(0.001)	(0.001)
profit					0.000*
					(0.000)
capital_labour					0.000
					(0.000)
un_sunk_cost					0.060
					(1.319)
regulation_realtime					0.032***
					(0.008)
african					-0.499***
					(0.157)
Constant	-0.014	0.035	-0.356***	-0.401***	2.600**
	(0.032)	(0.028)	(0.078)	(0.079)	(1.286)
Other control variables	NO	NO	NO	NO	YES
Industry Dummies	NO	NO	NO	NO	YES
Region Dummies	NO	NO	NO	NO	YES
Pseudo R-squared	0.003	0	0.011	0.014	0.200
Observations	2001	2001	2001	2001	2001
Log-Likelihood	-1381.4	-1386.2	-1371.2	-1366.7	-1109.2
Chi-Squared	9.3	0	29.4	38.2	348.4
Ind-Chi-Squared	N/A	N/A	N/A	N/A	9.4
Ind-P Value	N/A	N/A	N/A	N/A	0.224
Reg-Chi-Squared	N/A	N/A	N/A	N/A	193
Reg-P Value	N/A	N/A	N/A	N/A	0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$

Dependent variable is a dummy=1 if firm reported a positive bribe payment; 0 otherwise. Other control variables include: number of competitors (category); whether or not the government was a customer (binary); and the amount of money spent on external consultants (Naira). The coefficients on these variables are not statistically significant at the 10% level.

Table 3.12: Regressions On The Amount Of Bribe

	m0	m1	m10
bribereport	-42.928***	-42.958***	-35.826***
	(2.401)	(2.403)	(3.065)
infraserv	1.999	1.955	2.102
	(3.68)	(3.675)	(3.653)
import	-9.195		
	(11.654)		
export	4.416		
	(19.624)		
tax_percentage	0.024	0.023	-0.058
	(0.065)	(0.065)	(0.062)
trade		-5.737	-9.213
		(10.454)	(7.084)
eprofit			0.155***
			(0.035)
capital_labour			0.006**
			(0.003)
eun_sunk_cost			
un_sunk_cost			-5.058
			(59.839)
Industry Dummies			YES
Region Dummies			YES
Constant	47.195***	47.264***	44.830
	(5.431)	(5.43)	(48.844)
Other control variables	NO	NO	YES
F	65.652	82.062	.
Adjusted R-squared	0.075	0.076	0.527
Observations	1029	1029	1029
Ind-F			1.6
Ind-P Value			0.128
Reg-F			4
Reg-P Value			0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$

Dependent variable is bribe amount in '000 Naira. Other Control variables inlclude: number of competitors (categorical); government as a customer (dummy); amount of manager's time spent dealing with (state and local) government regulations per week; amount of Naira spent on external consultants per employee; and whether or not the owner is of African descent. These variables return insignificant coefficients in model 10.

The full specification (Model 12) shows no evidence for the (current or expected) profitability of a firm influencing whether or not it pays a bribe. The coefficient on profit (*profit*) and capital stock per employee (*capital_labour*) are both zero and insignificant in all models (Models 5 to 12, shown in Table A.24 in Appendix A.1 Section A.1.2). This means that firms with high and low profits are forced to pay bribes if they are required to interact with public officials relatively frequently. Moreover, firms reporting bribes and those reporting no bribes do not differ as a group in terms of profits (current or expected). The coefficients for (un)sunk cost (*un_sunk_cost*[34]) are negative and significant at the 1% level. This result gives support to the idea that firms consider outside options when faced with the possibility of paying bribes (Equation 3.2.0.3). It is also consistent with the idea that industries with a high level of sunk cost are more likely to be targeted by bribe demanding officials (e.g. the petroleum manufacturing industry).

The variables *gov_customer* and *regulation_realtime* measure whether or not the government is the principal buyer for the firms output and the amount of hours per week that senior managers spend dealing with requirements imposed by state/local and federal government regulations, respectively. These two variables help to increase the number of measures used to test the control rights theory. Both enter significantly into the estimation. Having 5 competitors or less also increases the probability that a firm will have to pay a bribe, as does a firm with an owner whose ethnic origin is not African. Expenditure on external consultants to deal with requirements imposed by federal government regulations is not a statistically significant predictor of bribe payment.

A likelihood ratio test for the coefficients on the bargaining measures being equal to zero fails to reject the null hypothesis of a zero effect of profit, capital stock and alternative return on the propensity to bribe. A test on the subsector dummies fails to reject the null of equality among the manufacturing sectors. Nevertheless, there appears to be regional differences in the propensity to bribe because a Wald test on equality of regional dummies rejects the null hypothesis of equality.

The coefficients in the probit model report the effect of an infinitesimal change in the explanatory variables (for continuous variables) or a discrete change from 0 to 1 (for dummy variables) on the standardised probit index associated with the paying of bribes. In order to understand the magnitude of the regressors on the probability of paying a bribe, marginal and impact effects must be computed using information from the standard normal probability density function and cumulative

[34]This variable measures the resale to replace value of capital stock (controlling for loss of value due to depreciation). A higher value indicates that the firm is able to retrieve a relatively higher share of its initial investment by selling. A low value means that a larger amount of the initial expenditure is gone (sunk), leaving the firm in a weaker bargaining position (less of an option to sell and relocate elsewhere.)

Table 3.13: Tobit Estimations On Bribe Amounts

	m0	m1	m10
model			
infraserv	10.027***	10.037***	6.373*
	(3.861)	(3.859)	(3.429)
import	-4.338		
	(13.769)		
export	-12.407		
	(24.374)		
trade		-9.123	-10.933
		(12.543)	(11.756)
tax_percentage	0.244***	0.245***	0.065
	(0.068)	(0.068)	(0.069)
eprofit			0.106***
			(0.028)
capital_labour			0.007***
			(0.002)
un_sunk_cost			14.372
			(66.191)
competitors_2			1.887
			(11.910)
competitors_3			18.610**
			(8.359)
competitors_4			13.238
			(8.665)
gov_customer			20.643
			(16.180)
regulation_realtime			1.522***
			(0.366)
eexternal_consultant_naira			0.221
			(0.240)
african			-13.315**
			(6.237)
Constant	-31.799***	-31.869***	61.734
	(7.603)	(7.615)	(56.849)
sigma	77.757***	77.752***	59.710***
	(10.586)	(10.586)	(6.396)
Industry Dummies			YES
Region Dummies			YES
Pseudo R-squared	0.0015	0.001	0.057
Observations	2001	2001	2001
Log-Likelihood	-6547.6	-6547.6	-6184.2
F	5.9	7.8	5.6
Ind-Chi-Squared			1.5
Ind-P Value			0.158
Reg-Chi-Squared			4
Reg-P Value			0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dependent variable is bribe amount in '000 Naira

Table 3.14: Probit Estimations On The Propensity To Bribe - NBS Data

	0	12
employee	-0.031	-0.074
	(0.09)	(0.102)
trade	0.907***	0.549
	(0.223)	(0.352)
foreign	-0.693***	-1.074***
	(0.229)	(0.311)
lage	-0.073	-0.073
	(0.099)	(0.116)
condb		-0.844*
		(0.475)
condg		-0.789**
		(0.357)
condh		1.067***
		(0.375)
condj		2.048***
		(0.348)
condk		-1.067**
		(0.534)
Constant	-0.751	-1.627***
	(0.654	(0.462)
Other control variables	YES	YES
State Dummies	YES	YES
Pseudo R-squared	0.15	0.383
Observations	287	287.000
Log-Likelihood	-137	-99
Chi-Squared	50.1	104.6
Reg-Chi-Squared	21.78	39.12
Reg-P Value	0.3	0.0

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dependent variable is a dummy variable equal to 1 if the firm reported paying a bribe; 0 otherwise. Other control variables, for both models, include: number of employees; and the log of firm age. The coefficients on these variables are not significant. Other control variables for model 12 include: condc; condd; conde; condf; and condi. The coefficients on these variables are not significant.

Table 3.15: Multinomial Logit Models On Different Types Of Bribery - NBS Data

	1	2	3	4
Dependent Variable	bribe_offer_official	bribe_demand_official	bribe_offer_foreign_official	bribe_demand_foreign_official
	very frequent	very frequent	very frequent	very frequent
trade	0.135	-0.001	-32.736***	-37.845***
	(0.8)	(0.93)	(2.714)	(1.349)
condb	0.041	-1.177	81.932***	33.557***
	(0.755)	(1.187)	(5.966)	(1.308)
condc	-1.486	-1.846	-162.073***	-13.930***
	(0.932)	(1.759)	(17.417)	(2.252)
condd	1.435*	2.292**	-176.051***	-71.154***
	(0.815)	(1.1)	(3.448)	(2.020)
conde	2.761***	1.453	210.280***	-36.502***
	(0.893)	(1.216)	(3.305)	(2.829)
condf	-0.622	0.007	-211.968***	-37.365***
	(0.827)	(1)	(4.256)	(1.213)
condg	-1.716*	-0.994	-67.154***	9.368***
	(0.995)	(1.526)	(3.763)	(1.674)
condh	-0.741	-2.141**	-76.649***	-38.522***
	(0.66)	(1.071)	(7.07)	(1.014)
condi	1.233	-2.391	54.870***	40.055***
	(1.36)	(1.673)	(7.193)	(1.476)
condj	3.023***	4.686***	-11.833***	-21.215***
	(0.756)	(1.103)	(1.914)	(2.833)
condk	-0.263	1.992	279.969***	38.685***
	(1.423)	(1.494)	(23.027)	(2.843)
Constant	-23.772	-28.020***	-1.908	-5.575
	(61.865)	(1.483)	(.)	(.)
State Dummies			YES	
Other control variables			YES	
Pseudo R-squared	0.335	0.375	0.483	0.424
Observations	284	282	248	245
Log-Likelihood	-229.4	-201.6	-87.4	-82.6

$*p < 0.10, **p < 0.05, ***p < 0.01$. Options were: never happens; not very frequent but not unusual; fairly frequent; and very frequent. Other control variables include: number of employees; foreign ownership (dummy); and the log of age of the firm.

distribution function, respectively. The marginal effect is: $\frac{\delta Pr[BribeDummy_i=1]}{\delta x_k} = \phi(\mathbf{x_i'}\beta) \times \beta_k$. The impact effect is: $\Delta = \Phi(\mathbf{x_i'}\beta + \gamma) - \Phi(\mathbf{x_i'}\beta)$.

Where ϕ represents the standard normal probability density function; β is the coefficient on the continuous variable and γ is the coefficient on the dummy variable. Table 3.16 shows the marginal and impact effects from the probit estimations. Dummy variables are labelled with a (d). The largest impact effect comes from selling mainly to the government. Having the government as the principal buyer of the firms output raises the probability of paying a bribe by 0.193. The public services variable is an index from 0 to 2. Results from model 12 show that firms who receive either water or electricity from the government[35] have an increased probability of paying bribes of magnitude 0.068 compared to firms who do not

[35] 24% of the sample report that they receive one of the two public services.

receive either of the public services[36]. A 1% increase in the precentage of sales that are reported for tax purposes corresponds to an increased chance of paying a bribe of 0.2% (0.002 probability points).

Once again the amount of profit or capital stock has no effect on the probability of paying a bribe. Having more than 5 competitors for the main product decreases the likelihood of paying a bribe by 0.067 probability points. Every hour (per week) spent dealing with government regulations increases the probability of paying a bribe by 0.013 probability points. On average, firms spent a mean of 3 hours per week dealing with government regulations. This translates to an increased probability of paying a bribe equal to 0.039 compared with firms who did not spend time dealing with government regulations. The values for (un)sunk cost ranged from -0.94 to -0.84. This translates into 39% to 43% of capital being transferrable. A decrease in this value of 1% translates to a decrease in the non-sunk cost variable of 0.025, which corresponds to a decreased probability of 0.0006.

3.8 Factors Influencing The Size Of Bribe Payment: Second Stage Analysis

Table 3.12 shows the results for the second stage of the two-part model. The dependent variable is reported bribe payments (in '000 Naira). Firms who reported their bribe payments in Naira paid significantly less in bribes per employee than those who reported their bribes as a percentage of sales. Results from the full specification (model 10) report that, on average, this difference is ₦36,000. Profits and alternative return both have their expected signs. Profit has a positive and significant effect on the amount of bribe paid, giving evidence for a company's "ability to pay" determining the amount of bribe that it pays, given that it pays a bribe. The coefficient on alternative return is negative and also significant, hinting that firms who are able to re-allocate their resources in another trade will pay lower bribes than businesses who are not so able. Similar to the first stage analysis, firms who have many competitors pay less in bribes and those that sell to the government pay more in bribes. A wald test for a zero coefficient on the control rights measures does not reject the null hypothesis that these parameters are zero at the 5 percent level. Together with the result from the probit estimations test statistic, this suggests that the factors that determine the occurrence of bribery (interaction with government) do not strongly influence the amount of bribe paid. Also, the factors that influence the amount paid (willingness/ability to pay) do not influence the propensity to pay. The models presented in Table 3.12 also seem to explain the data relatively well. The Adjusted R-Squared for model 10 is 0.527[37].

[36] 74% of the sample report that they receive both public services.

[37] The Adjusted R^2 value in Svensson (2000) was 0.21.

Table 3.16: Marginal And Impact Effects From First Stage Probit Estimations

	m1	m2	m3	m4	m12
infraserv	0.070***			0.068***	0.068***
	(0.023)			(0.023)	(0.026)
trade (d)		0.015		0.009	-0.000
		(0.086)		(0.086)	(0.095)
tax_percentage			0.002***	0.002***	0.001***
			(0.000)	(0.000)	(0.001)
profit					0.000*
					(0.000)
capital_labour					0.000
					(0.000)
un_sunk_cost					0.024
					(0.520)
competitors_2 (d)					0.020
					(0.133)
competitors_3 (d)					0.109
					(0.084)
competitors_4 (d)					0.058
					(0.086)
gov_customer (d)					0.121
					(0.105)
regulation_realtime					0.013***
					(0.003)
external_consultant_naira					0.000
					(0.000)
african (d)					-0.183***
					(0.052)

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dependent variable is a dummy equal to 1 if the firm reported paying a bribe; 0 otherwise. (d) for discrete change of dummy variable from 0 to 1.

3.9 Alternative Estimation: Tobit Analysis

Results from the Tobit estimations are presented in Table 3.13. The first two columns (models 0 and 1) only include the variables denoting the control rights hypothesis as explanatory variables. The first model disaggregates the trade variable into imports and exports. The coefficients on interactions with public officials ("infraserv") and percentage of sales reported for tax purposes ("tax_percentage") have their expected signs (both positive) and are statistically significant at the 1% significance level. This significance remains from model 0 to 9 for "tax_percentage" and from model 0 to 1 and 4 to 10 for "infraserv" (All Models are shown in Table A.27 in Appendix A.1 Section A.1.2). Therefore, the results point to the conclusion that the level of bribe payments is increasing in the receipt of public services and the reporting/revealing of sales for tax purposes, respectively.

Model 2 adds profit per employee to the estimation, the coefficient on this variable is positive and statistically significant at the 1% level. This is augmented by the capital stock measure ("capital_labour": the value of machinery, vehicles and equipment per worker) in model 3, which is positive and significant at the 5% level. These results suggest that visibility and ability to pay are positively related to the amount of bribe paid by the firm. Furthermore, they also suggest that the ability to pay now is more important than future ability to pay in determining the amount of bribe paid by firms. This is evident from the larger size and higher significance of the coefficient on profit compared to that of capital stock in all Tobit models. This tends to suggest that firms heavily discount the future when faced with a bribe demanding official. This is consistent with there being an increased level of uncertainty in a business environment where bribes are demanded.

Firms also seem to place a great emphasis on their outside options. Model 3 adds moveable capital per employee ("eun_sunk_cost"[38]) to the analysis which has a negative coefficient that is significant at the 1% level. Thus, companies with a relatively larger proportion of moveable capital pay significantly less in bribes, conforming with the bargaining theory. Model 5 introduces dummy variables denoting the number of competitors (1 competitor; 2 to 5 competitors; 6+ competitors) into the model (the reference category is: no competitors). Firms with more than 5 close competitors tend to pay less in bribes (significant at the 1% level), suggesting that a higher level of economic activity might decrease the amount of bribe to be paid (per firm) to a public official to keep him/her satisfied.

Having the government or government agencies as the principal buyer for one's product ("gov_customer") corresponds with a higher bribe payment (model 6; significant at the 10% level). This is coherent with the idea that firms pay bribes in order to secure public procurement contracts from the government. Firms engaging in international trade pay less in bribes than firms that do not. One reason for this could be that customs officials demand less in bribe payments than other types of public officials. The presence of many different types of government officials at Nigerian ports might increase the probability of being caught soliciting a bribe, this threat is likely to push down the bribe amounts for firms who import or export directly. When distinguishing between firms that import and those that export (model 0) it appears that importing firms pay more in bribes than exporting firms. This may be because exporting firms have the option of not exporting and cancelling an outstanding order, whereas an importing firm has probably paid for a cargo of goods. Also, the survey question asking whether or not the firm inported was in reference to inputs into the production process, it would seem plausible that a firm importing inputs might see the process as more urgent than one who was exporting due to hold up costs of production.

[38]This variable is constructed by dividing *un_sunk_cost* by the total number of employees

Specification Tests The likelihood function for the standard censored Tobit model can be expressed as the product of two likelihood functions; the likelihood functions for the truncated standard tobit model and for the probit model, respectively [Lin & Schmidt , 1984]. Recalling the likelihood function for the standard (censored) tobit model:

$$l = \prod_{y>0} \left[\frac{\phi[(y_i - \mathbf{x_i'}\beta)/\sigma]}{\sigma} \right] \prod \left[1 - \Phi[\frac{\mathbf{x_i'}\beta}{\sigma}] \right] \tag{3.9.0.22}$$

where $y = 0$ denotes the left-censored observations and $y > 0$ denotes the uncensored observations. Multiplying this expression by:

$$\prod_{y>0} \Phi \left[\frac{\mathbf{x_i'}\beta}{\sigma} \right] \prod_{y=0} \frac{1}{\Phi[\mathbf{x_i'}\beta/\sigma]} \tag{3.9.0.23}$$

is equivalent to multiplying by one, therefore the original expression remains unaltered. The following function is obtained:

$$l = \overbrace{\prod_{y>0} \left[\frac{\phi[y_i - \mathbf{x_i'}\beta/\sigma]}{\sigma\Phi[\mathbf{x_i'}\beta/\sigma]} \right]}^{l_{tt}} \times \underbrace{\prod_{y=0} \left[1 - \Phi[\frac{\mathbf{x_i'}\beta}{\sigma}] \right] \prod_{y>0} \Phi \frac{\mathbf{x_i'}\beta}{\sigma}}_{l_p} \tag{3.9.0.24}$$

where $l_{tt} = \prod_{y>0} \left[\frac{\phi[y_i - \mathbf{x_i'}\beta/\sigma]}{\sigma\Phi[\mathbf{x_i'}\beta/\sigma]} \right]$ is the likelihood function for the truncated tobit model and $l_p = \prod_{y=0} \left[1 - \Phi[\frac{\mathbf{x_i'}\beta}{\sigma}] \right] \prod_{y>0} \Phi \frac{\mathbf{x_i'}\beta}{\sigma}$ is the likelihood function for the probit model. Taking logs of equation 3.9.0.24 gives:

$$L = L_{TT} + L_P \tag{3.9.0.25}$$

where $L = log_e l$; $L_{TT} = log_e l_{tt}$; and $L_p = log_e l_p$. A likelihood ratio test (LRT) can be conducted to test for the restriction on parameters that the standard censored tobit model imposes. The unrestricted log-likelihood value is provided by equation 3.9.0.25 (reported in Table 3.17) and the restricted log-likelihood is reported in table 3.13. The LRT statistic is computed as: $LRT = -2[L_{CT} - (L_{TT} + L_P)]$, where L_{CT} is the log-likelihood value for the censored tobit model. This gives a formal test of the assumption that the factors that influence the (non)payment of a bribe retain the same sign when analysing their effect on the amount of bribe paid. Tables 3.17 and 3.18 report the results from the probit estimations and the truncated Tobit models, respectively. The LRT statistics are reported in Table 3.17 and results from all specifications do not reject the null hypothesis that the Tobit model is correctly specified.

Examining Differences Among Types Of Bribe Report As stated previously (section 3.6), firms were allowed to report their bribe payments either as a proportion of sales or in Naira value. This question structure lends itself to the analysis of whether or not there are any differences in the explanation of the type of response given by the firm. If there are some differences in the explanations for the responses, then the control rights & bargaining variables will be significantly different across the two group (Sales Percentage reporters & Naira reporters). This question can be explored using a multinomial logit model where:

$y_{ij} = 1$ if the i^{th} firm responds via the j^{th} alternative and $y_{ij} = 0$ otherwise, where $j = \{1, 2, 3\}$. Denoting $Prob[y_{ij} = 1] = p_{ij}$, then $p_{i1} + p_{i2} + p_{i3} = 1$; since the individual probabilities must add up to one.

The multinomial logit for this model can be expressed as: $p_{i1} = \frac{e^{\mathbf{x_i'}\beta_\mathbf{1}}}{e^{\mathbf{x_i'}\beta_\mathbf{1}}+e^{\mathbf{x_i'}\beta_\mathbf{2}}+e^{\mathbf{x_i'}\beta_\mathbf{3}}}$; $p_{i2} = \frac{e^{\mathbf{x_i'}\beta_\mathbf{2}}}{e^{\mathbf{x_i'}\beta_\mathbf{1}}+e^{\mathbf{x_i'}\beta_\mathbf{2}}+e^{\mathbf{x_i'}\beta_\mathbf{3}}}$; and $p_{i3} = \frac{e^{\mathbf{x_i'}\beta_\mathbf{3}}}{e^{\mathbf{x_i'}\beta_\mathbf{1}}+e^{\mathbf{x_i'}\beta_\mathbf{2}}+e^{\mathbf{x_i'}\beta_\mathbf{3}}}$

To ensure identification, the Thiel normalisation is used by setting $\beta_1 = 0$. After normalising the model becomes:

$$p_{i1} = \frac{1}{1 + e^{\mathbf{x_i'}\beta_\mathbf{2}} + e^{\mathbf{x_i'}\beta_\mathbf{3}}} \tag{3.9.0.26}$$

$$p_{i2} = \frac{e^{\mathbf{x_i'}\beta_\mathbf{2}}}{1 + e^{\mathbf{x_i'}\beta_\mathbf{2}} + e^{\mathbf{x_i'}\beta_\mathbf{3}}} \tag{3.9.0.27}$$

$$p_{i3} = \frac{e^{\mathbf{x_i'}\beta_\mathbf{3}}}{1 + e^{\mathbf{x_i'}\beta_\mathbf{2}} + e^{\mathbf{x_i'}\beta_\mathbf{3}}} \tag{3.9.0.28}$$

The log-odds ratios for these parameters are: $log_e[\frac{p_{i2}}{p_{i1}}] = \mathbf{x_i'}\beta_\mathbf{2}$; $log_e[\frac{p_{i3}}{p_{i1}}] = \mathbf{x_i'}\beta_\mathbf{3}$; and $log_e[\frac{p_{i3}}{p_{i2}}] = log_e[\frac{p_{i3}}{p_{i1}}] - log_e[\frac{p_{i2}}{p_{i1}}] = \mathbf{x_i'}\beta_\mathbf{3} - \mathbf{x_i'}\beta_\mathbf{2} = \mathbf{x_i'}[\beta_\mathbf{3} - \beta_\mathbf{2}]$.

Estimation of the parameters of the multinomial logit model is performed by using the following log-likelihood function:

$$L = \sum_{i}^{n}\sum_{1}^{3} y_{ij} log_e(p_{ij}) \tag{3.9.0.29}$$

where p_{ij} represents equations (3.9.0.26) to (3.9.0.28).

Marginal effects for a small change in one of the explanatory variables on the probability of one of the events occuring are given by (subscripts are dropped for convenience): $\frac{\partial p_j}{\partial x} = p_j[\beta_j - \sum_k \beta_k p_k] \quad for \quad j \neq k$ for the non-normalised categories (p_{i2} and p_{i3}) and $\frac{\partial p_1}{\partial x} = p_1[-\beta_2 p_2 - \beta_3 p_3]$ for the normalised category (p_{i1}).

Impact effects are given by: $\Delta_2 = \frac{e^{\mathbf{x_i'}\beta_\mathbf{2}+\gamma_2}}{1+e^{\mathbf{x_i'}\beta_\mathbf{2}+\gamma_2}+e^{\mathbf{x_i'}\beta_\mathbf{3}+\gamma_3}} - \frac{e^{\mathbf{x_i'}\beta_\mathbf{2}}}{1+e^{\mathbf{x_i'}\beta_\mathbf{2}}+e^{\mathbf{x_i'}\beta_\mathbf{2}}}$; and $\Delta_3 = \frac{e^{\mathbf{x_i'}\beta_\mathbf{3}+\gamma_3}}{1+e^{\mathbf{x_i'}\beta_\mathbf{2}+\gamma_2}+e^{\mathbf{x_i'}\beta_\mathbf{3}+\gamma_3}} - \frac{e^{\mathbf{x_i'}\beta_\mathbf{3}}}{1+e^{\mathbf{x_i'}\beta_\mathbf{2}}+e^{\mathbf{x_i'}\beta_\mathbf{2}}}$ where γ is the coefficient on the dummy variable.

Table 3.17: Probit Estimations On All Firms

	p0	p1	p10
infraserv	0.169***	0.170***	0.182***
	(0.057)	(0.057)	(0.067)
import	0.174		
	(0.262)		
export	-0.183		
	(0.361)		
tax_percentage	0.006***	0.006***	0.004***
	(0.001)	(0.001)	(0.001)
trade		0.023	0.010
		(0.216)	(0.237)
eprofit			-0.000
			(0.000)
capital_labour			0.000
			(0.000)
eun_sunk_cost			
competitors_2			0.054
			(0.341)
competitors_3			0.294
			(0.223)
competitors_4			0.158
			(0.216)
gov_customer			0.337
			(0.293)
regulation_realtime			0.032***
			(0.008)
eexternal _consultant _naira			0.001
			(0.002)
african			-0.507***
			(0.157)
un_sunk_cost			-0.025
			(1.317)
Industry Dummies			YES
Region Dummies			YES
Constant	-0.398***	-0.401***	2.572**
	(0.079	(0.079)	(1.284)
Pseudo R-squared	0.014	0.014	0.199
Observations	2001	2001	2001
Log-Likelihood	-1366.4	-1366.7	-1110.8
L	-7092.4	-7092.8	-6468.5
Lrt	-1089.6	-1090.4	-568.6

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dependent Variable: Dummy=1 if bribe paid, 0 otherwise. L: the sum of the log-likelihood for the truncated tobit model and the log-likelihood for the probit model. This is the log-likelihood for the unrestricted model in the tobit specification test. LRT: the likelihood ratio test statistic for the tobit model specification test.

Table 3.18: Truncated Tobit Estimations For Bribing Firms Only

	tt0	tt1	tt10
infraserv	2.228	2.161	2.589
	(3.81)	(3.807)	(3.682)
import	-18.347		
	(12.432)		
export	-0.171		
	(22.321)		
tax_percentage	-0.086	-0.087	-0.122*
	(0.068)	(0.068)	(0.063)
trade		-14.815	-12.515
		(11.35)	(7.747)
eprofit			0.155***
			(0.034)
capital_labour			0.007**
			(0.003)
eun_sunk_cost			
competitors_2			-0.217
			(9.617)
competitors_3			5.517
			(5.931)
competitors_4			6.999
			(6.541)
gov_customer			15.720
			(14.424)
regulation_realtime			0.081
			(0.456)
eexternal _consultant _naira			0.720
			(0.475)
african			4.375
			(3.978)
un_sunk_cost			20.741
			(60.197)
Industry Dummies			YES
Region Dummies			YES
Constant	45.806***	45.892***	63.599
	(5.598)	(5.602)	(49.687)
sigma	63.161***	63.170***	44.160***
	(10.604)	(10.603)	(4.985)
Pseudo R-squared	0	0	0.065
Observations	1029	1029	1029
Log-Likelihood	-5726	-5726.1	-5357.7
F	1	1.1	14.3

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dependent Variable is Bribe ('000 Naira) per Employee

Results Taking the exponential of the reported multinomial logit coefficients provides the relative risk ratios for the estimation:

$$\frac{p_{ij}}{p_{i1}} = e^{\mathbf{x_i'}\beta_\mathbf{j}} \tag{3.9.0.30}$$

where $j = 2, 3$. This gives the proportionate change in the relative risk of choosing alternative j rather than "no bribe" from a unit increase in an explanatory variable. The relative risk ratios for the estimation are reported in Table 3.19. This table uses the same explanatory variables as model 12 of the probit model in table 3.14. As stated above, firms had the option of reporting no bribes, positive bribes as a percentage of sales, or positive bribes in Naira. The base group for the reported results is the non-payment of bribes.

Results indicate that firms who report bribes in Naira are more responsive to the factors affecting bribery, compared to the base group of non-bribers, than firms who report in sales percentage. The regressors in model 12 are jointly significant (chi-squared value of 126197) and the Pseudo R-squared is larger than that of more sparse models. The coefficients on *tax_percentage*; *regulation_realtime*; and *african* are all jointly significant at the 1% level, those on *infraserv* and *external_consultant_naira* are significant at the 5% level; while the coefficients on *profit* and *capital_labour* are significant at the 10% level. All coefficients for jointly significant variables remain the same between the two groups within each model[39].

In this table, coefficients that are less than 1 imply that a unit change in the explanatory variables decrease the probability of choosing that particular reporting style. So in model 12, a one unit increase in *regulation_realtime*, corresponding to an extra hour spent dealing with government regulations, leads to the relative odds of reporting a bribe in sales rather than reporting a zero bribe that are 1.054 times what they were before the change, therefore the relative odds have increased. Consistent with previous results, changes in profits do not change the relative odds of reporting bribes in either Naira or sales relative to reporting no bribe at all. Acquiring access to another public good (water or electricity) increases the relative risk of reporting a bribe (in either Naira of sales), compared to not reporting, that is greater than any other explanatory variable. *regulation_realtime* has the second largest relative risk ratio that is greater than 1, and *tax_percentage* has the third. This implies that an extra hour spent dealing with government regulations is associated with an increased relative risk of paying a bribe that is larger than a 1 percentage point increase in the percentage of sales reported for tax purposes.

[39]Excluding *external_consultant_naira* and *capital_labour*, which, in model 12, have a zero coefficient for sales reporters and a negative and significant coefficient for Naira reporters

Table 3.19: Relative Risk Ratios For The Multinomial Logit Estimations On Bribe Reporting Behaviour

	ml12	
	Sales Reporter	Naira Reporter
infraserv	1.338**	1.106
	(0.152)	(0.209)
tax_percentage	1.004*	1.015***
	(0.002)	(0.004)
profit	1.000**	1.000
	(0.000)	(0.000)
capital_labour	1.000	1.000*
	(0.000)	(0.000)
regulation_realtime	1.054***	1.066***
	(0.016)	(0.023)
external_consultant_naira	1.000	0.996**
	(0.000)	(0.002)
african	0.436***	0.550
	(0.114)	(0.268)
Industry Dummies	YES	YES
Region Dummies	YES	YES
Constant	518.315***	0.272
	-1243.178	(0.975)
Other control variables	YES	YES
Pseudo R-squared	0.226	
Observations	2001	
Log-Likelihood	-1493.4	
Chi-Squared	126197.5	

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dependent Variable={No bribe;Bribe(Sales %); Bribe(₦)}. Reference group: Non-bribers. Other control variables include: trade (dummy), capital mobility; the number of competitors (category); and whether or not the government was the principal buyer of the firm's output (dummy). The coefficients on these variables are not statistically significant.

The Independence Of Irrelevant Alternatives Assumption One potential flaw with using the multinomial logit (MNL) model in this analysis is that it is equivalent to a series of pairwise logit models. Following from this, it assumes that the probability under consideration is unaffected by all other alternative choices. I.e. the odds of choosing to report bribes in Naira rather than no bribes does not depend on how many other options there are in total. This assumption reduces the MNL model to a series of binary choice logit models (this can be seen in Equation 3.9.0.30, the conditional probability of one event occuring , e.g. Naira, does not depend on the other probabilities). This is called the "independence of irrelevant alternatives" assumption (IIA) and is potentially problematic when some choices are close substitutes for one another. In the current case, there seems to be a very close substitutability between reporting a bribe in Naira and reporting a bribe as a percentage of total sales. These two options are not complementary because no firms reported bribes both in terms of Naira and percentage of sales. In order for the MNL model to be valid, the IIA assumption must be tested. To do so, this study follows the convention; the procedure and results of the IIA testing are explained in Appendix A.1 Section A.1.4.

3.10 Results From The Corruption Survey

This section augments the previous results by including data from the NBS survey into the analysis. The previous analysis has two potential limitations: Firstly, the data on bribery does not specify one particular act for which the bribe was paid but a host of acts. Firms were asked whether payments were made with regard to "customs, taxes, licenses, regulations, services etc". One the one hand, asking a question in this manner might attract a higher estimate of the proportion of bribing firms i.e. asking if a firm paid a bribe for a number of activities might increase the reporting of bribery. On the other hand, this measure does not distinguish between the bribes paid for different activities. Therefore, any conclusions based on this data might give a biased view of the business environment in Nigeria: a large proportion of firms might have paid a bribe in order to avoid punishment for an offence committed, rather than to bypass cumbersome regulation. In order to get a better view of the nature of bribe payments, the NBS data asks specific questions concerning what the bribes were paid for. This data allows for an analysis into the processes which attract the highest proportion of bribes.

Another potential problem with the previous data set is that it does not distinguish between the offfering of bribes (by firms) and the demanding of bribes (by public officials). The analysis in section 3.2 suggests that the supply of bribes is determined by the probability that an official will ask for a bribe, and the firms willingness to pay. If a bribe is paid it could be the case that a bribe was first offered by the firm, or that a bribe was initially demanded by the bureaucrat. The previous analysis suggests that a bribe is demanded by an official and that the firm either pays of refuses. The NBS data distinguishes between bribes being demanded by public officials and bribes being offered to public officials. This allows for an investigation into the nature of causality surrounding bribe payments in Nigeria.

The current section looks at the factors determining bribery by introducing a disaggregated set of variables which represent meetings with public officials. Tables A.1 to A.11 (in Appendix Section A.1) use these variables along with a disaggregated set of bribery dummy variables to address the first potential problem mentioned above of different types of bribery. It presents the results of the Bivariate Probit Analysis of the bribery of public officials that distinguishes between different types of public officials. The final part of this section (Section 3.10) uses both sets of variables to address the second problem of causality in the bribe transaction.

The Propensity to Bribe Table 3.14 presents results of the probit analysis on bribe payments. The dependent variable for the models in this table is a dummy variable equal to one if the firm admitted to paying any bribe, and 0 if the firm did not. Results show that involvement in international trade is associated with an increased probability of paying a bribe. The coefficient on *trade* is positive and statisticaly significant at the 1% level. The coefficient on *foreign* is negative and also significant at the 1% level, suggesting that amongst these firms, foreign owned businesses were less likely to pay bribes than Nigerian ones. The number of employees and age of a firm both have negative coefficients but neither are significant at the 10% level.

All of the coefficients for variables relating to contact with public officials (conda-condk) are positive when entered seperately into the model . The variables denoting traffic offences (condj); clearing goods through customs; police investigations (condi); vehicle registrations (condh); getting clearance for environmental or sanitary regulations (condf); and obtaining business licenses and permits (condd) all enter significantly into their respective models (in decreasing order of size). Hence, traffic offences, customs process and police investigations are the processes which seem to make a firm most likely to pay a bribe. The parameters of the partial effects models are jointly significant and have chi-squared values all above 53. There does not appear to be much evidence in favour of differences between regions in the payment of bribes: a Wald test for equality of the state dummies fails to reject the null of equality in all but one specification out of the partial effects models [40].

In the full model (model 12) which includes all of the public-official variables, only the variables for traffic offences and vehicle registrations enter positively and statistically significantly (both at the 1% level) into the model. The parameters of the model are jointly significant and the model also rejects the null hypothesis of equality of state dummies. The model has the highest Pseudo R-Squared value (0.384).

Instigator Of The Bribe Transaction Finally, table 3.15 helps to shine light on the instigator of the typical bribery transaction. This is done by using a series of categorical response questions which ask the firms how frequently bribes are demanded and offered to public officials and foreign public officials. Firms were given a choice of responses: never happens; not very frequent but not unusual; fairly frequent; and very frequent. The base category used in Table 3.15 is “never happens”.

The only variable that seems to have a strong effect across all models is the indicator variable representing coming into contact with public officials because of a traffic offence. This variable seems to be a good predictor of a bribe being offered to and demanded from a domestic public official, respectively. Unsurprisingly, this variable is negatively related to a bribe bribe being offered to and demanded from

[40]The model where the null of equality is rejected (model 10) has the best fit out of all partial effects models (Pseudo R-squared of 0.314).

a foreign public official; this is because traffic offences are dealt with by domestic officials and not foreign ones.

Starting with model 1, committing a traffic offence is associated with an increased log-odds of reporting that firms offer bribes to public officials very frequently relative to reporting that this never happens. Also, procuring goods and services from private companies increases the log relative-risk of saying that companies offer bribes to officials on a very frequent basis relative to that of saying that this never happens. This result is statistically significant at the 1% level. This increased risk holds for the "fairly frequent" and "not very frequent but not unusual" categories as well.

In addition to an increased risk of offering an official an informal payment, being involved in a traffic offence increases the log of the relative-risk of reporting that officials demand bribes very often relative to not at all (model 2). This result also holds for the other two categories. It therefore seems that traffic offences attract bribe demands from public officials and propel bribe offers from firms. The obtaining of a business licence or permit (*condd*) is also associated with an increased relative risk of a bribe being demanded relative to no bribe being demanded. Models 3 and 4 show a positive association between the acquiring of road worthy certificates and the log relative-risk of both being demanded for a bribe by and offering a bribe to a foreign public official. This result also holds for residence work permits, police investigations and contact with the courts.

3.11 Conclusions

Bribery at the firm level is a relatively unexplored topic. The lack of research in this area is partly due to data issues. This chapter sought to uncover some of the mechanisms at work when a firm manager pays a bribe to a government official by using firm level data that includes information on informal gift giving from firms to government employees. Information from two datasets of Nigerian firms was applied to an established conceptual framework provided by Svensson [2003] and results were significant and supportive of the previous literature.

The control rights and bargaining hypotheses were tested using a number of empirical models including: a tobit model; the Heckman Two-Step model; and the two-part model. Due to the presence of heteroscedasticity; absence of selection; and preference for increased flexibility, the Two-Part model was used as the preferred model. This model allowed for the independence of the determinants of the incidence of bribery and the magnitude of bribery (the Tobit does not); does not require an exclusion restriction (the Heckman does); and is the most robust simple estimator in the presence of multicollinearity [Puhani , 2000]. The models controlled for regional and industry fixed effects.

With suitable data, a bivariate probit model with sample selection was used to simultaneously model the propensity to come into contact with a government employee and the propensity to bribe that government employee conditional on coming into contact with them. This enabled the investigation to test for selectivity in the meeting and the bribing of officials. This amounted to a relaxing of one of the assumptions of the conceptual framework.

A major contribution of this work is the use of data that distinguishes between bribe payments made to different types of government official. This is the first study to do so at the firm level and it allowed for the examination of which types of bribery are most prevalent in the economy. Another significant contribution was the use of data concerning how likely it is for a firm to offer a bribe to a public official; and for a public official to demand a bribe from a firm. This information was available concerning domestic and foreign public officials. This is important because it relaxes some of the assumptions of the conceptual framework; it also provides a framework that can possibly be extended and investigated by the experimental literature.

The findings revealed evidence for both the control rights hypothesis and the bargaining hypothesis amongst manufacturing companies in Nigeria. Despite the potential sensitivity of the topic, various methods have been used in order to extract honest responses about bribery from firms. These responses lend to some analysis on the nature of bribe payments amongst Nigerian firms. Questions asking about the behaviour of other firms help to remove the stigma from admitting to doing a sensitive or illegal act.

The incidence of bribery seems to be positively related to coming into contact with government officials. Results also suggest that the magnitude of bribe is determined by factors representing a firm's ability to pay and outside options. These findings are robust to different econometric specifications. Results point to the notion that firms pay bribes because they desire items and/or services in order to carry out their business, which sometimes require them to pay informal gifts in order to either receive, or speed up the process of receiving them [Leite & Weidemann , 2002]. In the case of Nigeria, 65% of manufacturing firms state that the government's interpretation of laws and regulations are consistent and predictable. 51% of firms report a positive bribe amount that is required in order to speed the process of regulations.

By distinguishing between the different purposes for which a bribe was paid and who brought up the issue of a bribe, this investigation was able to find out which activities were more likely to attract bribes and to assess the direction of causality in the bribe transaction. The data presented in the preceeding analysis builds on the literature that analyses specific reasons for such payments.

Variables denoting meetings with public officials were found to be positively related with the propensity to pay a bribe and were also related (although not as strongly) to the amount of bribe paid. Firm specific characteristics such as profit and capital stock were found to be relatively good indicators of the level of bribe paid. Further work can build on this study by investigating whether informal payments have a significant effect on the future profitability and competitiveness of companies. This will allow for a better interpretation of the effect that expected future profits has on the level of bribe paid. Work that investigates the effect of competitors paying bribes would also add to the analysis of the supply of bribes.

The inference from the ES dataset relied on indrectly asking firm managers about their bribing habits. Much of the inference about firms bribing was based on their responses to these indirect questions. The next chapter calls the validity of such inference into question by testing the effectiveness of indirect questioning in getting honest responses from sensitive questions compared to another widely used method of asking sensitive questions.

4 A Comparison Of Randomised Response and Indirect Questioning Methods In Measuring Corruption And Tax Evasion: A Study Of Companies In Nigeria

Abstract

When using business surveys to study bribery, one must overcome the problem of acquiring valid data on such acts. Firms must be provided with an incentive to reveal their true behaviour if the data is to be trusted. This might be achieved by increasing the benefit accrued from truth-telling; or by reducing the perceived cost of truth-telling. Different mechanisms have been used to do this; these include the randomised response technique; and indirect questioning methods. These methods shield interviewees from being identified as guilty whilst providing some information about the level of guilt within the sample. This paper assesses the usefulness of these techniques in obtaining information about corruption and tax evasion amongst companies in Nigeria. Results from the randomised response procedure are compared with indirect questioning methods that are used to get truthful responses to sensitive questions. Through a model that allows for false-reporting this study calculates a lower bound estimate of the rate of truth-telling/lying. Results from previous studies using data from other countries are compared with the results from the Nigerian sample. This chapter finds that whilst the randomised response technique provides an incentive to tell the truth; many firms choose not to do so. Furthermore, asking indirect questions about tax evasion and corruption seems to generate more honest answers than using the randomised response technique. Companies in Romania seem more willing to admit to guilt than companies in Nigeria.

4.1 Measuring Corruption At The Micro Level

The main result of the previous chapter used indirect questions asked to firm managers about their committing of bribery. The firms were not asked about their own behaviour but about the behaviour of similar establishments. Nonetheless, their responses were interpreted as describing their own behaviour. The justification for this interpretation of the data comes from the results of the false consensus effect: that people who engage in socially undesirable behaviour are more likely to over-estimate the prevalence of such behaviour amongst their peers. Thus, firms engaging in bribery are more likely to over estimate the prevalence of bribery by similar establishments. The current chapter seeks to compare the effectiveness of indirect questioning against an alternative method of asking interviewees for sensitive information.

Using survey data to ask firm managers about corruption and tax evasion can generate bias due to misreporting. Managers are likely to underreport the amount of illegal activities they engage in due to either psychic costs or the risk of their responses being leaked to authorities [Tourangeau, Rips & Rasinski , 2000]. Techniques have been developed with the aim of getting honest responses from sensitive survey questions; these include: the assurance of the interviewee's anonymity; and indirect questioning. Both of these methods seek to acquire information about the respondent's behaviour whilst protecting them from the costs of confessing to an illegal or socially tabooed act [Lensvelt-Mulders et al. , 2005b]. Making the responses to a survey anonymous reduces the chances that any individual transgressor is punished; indirect questioning involves asking questions about similar firms or "firms like this one" rather than asking direct questions about the firm. This allows the firm to admit that an illegal act is common amongst similar firms without running the risk of being identified as guilty.

Another method that has been used in empirical research is the randomised response (RR) technique [Warner , 1965]. This procedure aims to get information about people's behaviour by partly removing the interviewee's true behaviour from their response. The method that is used in the current dissertation is equivalent to posing an innocuous question alongside the sensitive question and having the interviewee answer only one of the questions: either the sensitive question; or the non-sensitive question. The question that the interviewee is required to answer is determined by the occurrence of a random event (a heads on a coin toss), the result of which is known only to the interviewee, but the probability of which is known to the interviewer. The idea behind this procedure is that interviewees would be more willing to tell the truth if the interviewer or the users of the data do not know which question is answered. If uncertainty about which question was answered causes interviewees to be honest in their responses, then it is trivial to calculate unbiased estimates of the prevalence of sensitive behaviour [Kraay &

Murrell , 2013].

In the case of the dichotomous response to a question such as: "Have you failed to pay some of your taxes in the past 12 months?". The response (yes/no) is attached to a probability distribution that is independent of the interviewee's guilt, for example, receiving a heads from a coin toss. The probability of the event occurring is known to the interviewer, and the interviewee is instructed to answer "yes" based on the outcome of the random event. Under these conditions a "yes" is answered, according to a known probability, regardless of whether or not the person in question is guilty or not. The interviewer observes the response ("yes" or "no") but does not know the true status of the interviewee or the result of the random event. The assurance that a "yes" response does not necessarily imply guilt on the part of the interviewee is meant to reduce the incentive for the interviewee to tell a lie; since he/she could just be following instruction. Using the proportion of "yes" responses and the known probability of observing a "yes" response, the interviewer is able to get an estimate of the proportion of interviewees who are guilty of the act in question. This technique has been used in the economics, sociology, and psychology literature to estimate the prevalence of sensitive behaviour [Locander, Sidman & Bradburn , 1976, Lensvelt-Mulders et al. , 2005b] and identify indicators of sensitive and illegal behaviour [St. John; Keane; Edward-Jones; Jones; Yarnell and Jones , 2011, Kerkvliet , 1994a,b].

The current study uses the RR technique and indirect questioning methods to analyse the prevalence of corruption and tax avoidance amongst businesses in Nigeria. The methods are tested against each other in order to see which one gets more honest responses from interviewees. In testing these methods this study uses the "more is better" assumption [Tourangeau & Yan , 2007] which states that the method which yields a higher estimate of the prevalence of the sensitive behaviour is the more accurate method. This assumption applies only to questions that are generally underreported. For questions concerning socially desirable behaviour the opposite assumption, "less is better", applies.

The chapter also discusses the use of the RR technique in dealing with the response bias that arises out of asking questions about sensitive topics. Through the use of the RR technique this study is able to get an estimate of the level of misreporting (as done in previous literature [Azfar & Murrell , 2009, Clausen, Kraay & Murrell , 2010]). The analysis is applied to Nigeria.[41]

It has been relatively difficult to collect accurate data on acts such as tax evasion and corruption because they are usually carried out in secret due to their being illegal [Otusanya , 2011]. Furthermore, firms have been deemed to be unlikely

[41] This study builds on multiple strands of literature, including: misclassification error in a binary dependent variable; response bias when answering sensitive questions; improving the efficiency of the RR technique; testing the effectiveness of indirect questioning versus randomised response questioning; and estimating a likelihood function for data subject to RR.

to admit to their committing an illegal act. Empirical work on topics involving crime have sometimes generated new techniques in order to measure, create proxies for, or instrument for the variables of interest [Golden & Picci , 2005, Fisman & Miguel , 2007, Reinikka & Svensson , 2005]. Faced with the problem of quantifying an action that is socially undesirable, researchers have created indicators of such phenomena from observables that are related to these acts. Golden & Picci [2005] use public data about the amount of funds set aside for the building of public infrastructure in Italy and compare this to independent estimates of the costs of the public projects to estimate the amount of funds being stolen. They use this as a measure of corruption. Fisman & Miguel [2007] use the number of parking violations committed by foreign diplomats in New York as an international corruption index. Reinikka & Svensson [2005] use a natural experiment where local residents are provided with information about the level of funding given to Ugandan schools to gauge how governance and accountability reacts to the public provision of information. These techniques have allowed researchers to examine the rate at which funds are being unlawfully expropriated for agents' private use; and also how a culture of corruption and lawlessness can manifest itself outside the borders of a particular country [Fisman & Miguel , 2007, Barr & Serra , 2010].

Another data collection strategy has been to acquire information sourced in the public domain. These include: parliamentary reports; court proceedings; newspaper reports; and papers from regulatory bodies [Otusanya , 2011]. Each source has its own relative strengths and weaknesses. Most suffer from at least one form of selection problem. In the case of judicial proceedings the selection bias occurs in at least 2 stages. Firstly, individuals/companies must be selected in order to be tried and prosecuted; and not all perpetrators of the act might be selected. Secondly, the judiciary decides whether or not the information of the proceedings are made publicly available. Another data collection strategy is to randomly select individuals from a telephone directory [Abeler, Becker & Falk , 2012]. However, this also suffers from the selection bias and nonresponse bias [Porcano , 1988]. The current chapter does not use such information directly but uses firm-level surveys to get information on tax evasion and corruption. The argument is put forward that valid information can be acquired by companies on the nature of their dealings through the use of indirect questions. These include questions about firms that are similar to the one in question rather than direct questions about the firm itself. Compared to RR questions, these seem to generate a higher rate of admitting ones guilt.

The current study contributes to the body of knowledge by introducing a formal analysis of the responses to a series of randomised response questions with reticence; and by introducing new approaches to survey research. Corruption estimates are generated via the randomised response method, these are compared to estimates from the indirect questioning method (Section 4.5.1) . The distribution of responses to the set of randomised response questions are analysed in order to identify which firm managers are reticent and which ones are not (Section 4.5.2). The likelihood of observing the distribution of responses is modelled using a simple binary response framework, this is used to derive estimates of reticence and guilt from the sample of firms (Section 4.5.3). This method can be used to derive estimates of reticence and guilt from any sample of observations that use a set of RR questions. The RR questions were asked in a given order. The current study looks at whether or not there is a trend in the responses to a series of RR questions, it does this by observing the predictability of the response to the final RR question based on the responses to the previous RR questions (Section 4.5.4). This discussion is followed by an investigation on the order of the resposnes to all RR questions, common trends in the responses to the questions are identified (Section 4.5.5). The indirect questions were asked about firm that are similar to the firm being interviewed, the current study seeks to find out how much information the response to these questions might contain by investigating the responses of firms in the same sector-size-location cell (Section 4.5.6).

The rest of the chapter is organised as follows: The next section discusses the formation of the IQ and RR techniques. Section 4.3 discusses the data used in this study to investigate firm behaviour and illegal practices amongst companies in Nigeria. Section 4.4 specifies the methodology that is used to compare the two techniques. Section 4.5 analyses the results. Section 4.6 concludes and proposes possible extensions to the study.

4.2 Survey Research Methods And A Formal Description Of The Randomised Response Model

Firm-level data on production processes, pricing, expectations and appraisals of the business environment have been collected in many countries over many decades. These surveys range from monthly studies like the ones conducted by Office For National Statistics (United Kingdom) to less frequent ones such as the World Business Environment Survey (conducted by the World Bank). Historically, business surveys have been used to provide information for macroeconomic analysis, such as generating indicators for the business cycle. Recent progress in microeconometric methods have made it possible to use business surveys to test the microeconomic theory of the firm. For example, the use of business surveys has allowed for the empirical measurement of firm production functions and estimates of efficiency [Greene , 2008, Pham, Dao & Reilly , 2010].

When using surveys to come to conclusions about the business environment it is important for the data to be measured as accurately as possible [Good &

Hardin , 2006]. In order to do this, questions must be constructed effectively. Inaccurate responses can have a significant impact on a study via inaccurate point estimates and biased regression coefficients [Hausman, Abrevaya, Scott-Morton , 1998, Bound, Brown & Mathiowetz , 2001]. The importance of question structure is even more relevant when asking sensitive questions. Sensitive questions deal with behaviour that, when answered truthfully, is seen by society as illegal or undesirable, or is seen by the interviewee as an invasion of their privacy [Iarossi , 2006]. Some examples include questions concerning the payment of bribes, tax evasion, and the breaking of laws. Such questions tend to have a relatively higher risk of response error and non-response than non-sensitive questions. Respondents are likely to distort their answers in order to present themselves in the most socially positive way. Also, some their interviewees might refuse to answer the sensitive questions. If the pattern of non-responses is not random, as is likely, then the results will be biased; those with the most to hide might be the least likely to give a response. Another potential source or error that applies to sensitive questions is the bias generating by the interviewer's sponsoring agency. When a survey is conducted by a government agency sensitive questions on sales, corruption and the perception of corruption tend to be underreported compared to when the questionnaire is conducted by a private local survey company [Iarossi , 2006].

Two major forces are thought to be at work to distort the answer to a sensitive question. These are: the desire to avoid an answer that could pose a threat to the respondent; and the desire of the interviewee not to look bad [Fowler , 1995]. Different methods have been used to reduce the impact of these forces and get more accurate responses to sensitive questions. One method is to change the format of the question. This can be achieved by adjusting the level of detail required in the answers to the questions. Instead of asking for answers to be made in absolute values it might be easier for the interviewee to respond in percentages (e.g. the amount of bribe paid as a percentage of total sales) or categories (e.g. bribes are paid: not at all; sometimes; frequently). However, although changing the format of a sensitive question can improve the response rate, a potential downfall of this strategy is the loss in accuracy that might occur from the responses to the adjusted question. Whilst a direct sensitive question is subject to the sources of error mentioned above: underreporting, nonresponse and survey agency effects; a question that asks for a bribe amount as a percentage of sales might be subject to misreporting of bribes, misreporting of sales and miscalculation of the percentage [Clarke , 2011].

Another strategy used to reduce the bias that arises from asking sensitive questions is to adjust the length of the question. The length of a question can serve to reduce the threatening nature of a question. In addition to this, experiments have shown that longer questions generate more accurate answers when asking about sensitive topics concerning behaviour [Sudnam & Bradburn , 1974, Peterson , 2000]. A similar strategy is to reduce the perceived sensitivity of the topic by explaining the various reasons why somebody might engage in the behaviour in question. This might help to reduce the extent to which the interviewees feel their response will be viewed in a negative fashion. Another way in which this can be done is to use words that imply that the same behaviour is done by their peers. This may assist in making the interviewee feel less stigma in admitting to guilt. Also, the interviewer can ensure the interviewee that their answers will be made confidential and that none of the responses will be traceable back to them.

Sometimes a question is perceived to be sensitive because the interviewee does not understand why it is being asked in the questionnaire; perhaps they do not see the connection between the question and the purpose of the survey or do not see the importance of their answer. The interviewer can help to reduce this sensitivity by explaining to the interviewee the reason why the question is included in the survey. The sensitivity can also be reduced by putting the topic in a list of less sensitive topics [Plateck, Pierre-Pierre & Stevens , 1985].

The location of the sensitive question within the questionnaire can also be of importance when trying to get accurate answers. Placing the question where it is less sensitive, e.g. where the topic being questioned is apt for a sensitive question, has been advised [Iarossi , 2006]. The general consensus seems to be that interviewers should try to gain trust[42] from the interviewee before asking a potentially threatening questioning, therefore, sensitive questions should not come at the beginning of a questionnaire. It is also thought to be good practice to introduce the sensitive questions with a series of warm-up questions [Warwick & Lininger , 1975], for example by placing multiple questions on corruption in different subsections of a questionnaire [Svensson , 2003].

Another method that can be used to deal with sensitive topics is the randomised response (henceforth, RR) technique. With this technique the interviewee is presented with 2 questions: a sensitive question and a non-sensitive question[43]. The process of choosing which question to answer relies on a randomising device (e.g. a coin). The interviewee chooses which question to answer according to a predetermined probability (e.g. a coin toss. Heads: sensitive question; tails: non-sensitive question). The interviewer observes the interviewee's answer but does not see the result of the randomising procedure; so they do not know which question the inter-

[42]The potential role of trust in getting honest answers to sensitive survey questions is investigated in Chapter 5.

[43]The second question might be related or unrelated to the (sensitive) question of interest. For example, if the sensitive question is: "Have you paid a bribe this year?" the second question might be related: "Have you refrained from paying a bribe this year?" or unrelated: "Does your birthday land in August?". The initial RR framework [Warner , 1965] used the first type of question (the related question) as the second question.

viewee is answering. This is meant to reduce the perceived threat of the question and encourage honesty on the part of the interviewee. One can calculate the sample value of the sensitive question by using a weighted average of the probability of selecting each question [Moser & Kalton , 1971].

The Randomised Response Technique The RR technique increases the anonymity of the interviewee by attaching their response to a chance event. There is a chance that they are answering the sensitive question, and there is a chance that they are not. This uncertainty is meant to lower the threat associated with answering sensitive questions and encourage the person to respond with the truth. The RR method was initially developed by Warner [1965] to get estimates of the percentage of a sample that was guilty of a sensitive act. More recent papers have used the procedure to study the determinants of sensitive behaviour [St. John; Keane; Edward-Jones; Jones; Yarnell and Jones , 2011], and also to see if there is a link between lying about sensitive questions asked via RR and misreporting other questions [Azfar & Murrell , 2009, Clausen, Kraay & Murrell , 2010].

Asking a room full of 100 firm managers to raise their hands if they had paid a bribe in the past year might not achieve many raised hands. However, if one asks them to all flip a coin and after the coin toss to raise their hands if they had paid a bribe in the past year *or* if they had got a head on the coin toss, one might expect the number of raised hands to be more than the direct question approach. Furthermore, if, after using the RR approach, 64 people raised their hands, it can be expected that roughly 50 people got a heads and raised their hands regardless of whether or not they had paid a bribe. The other 14 hand raisers would suggest that 14 out of the (approximately) 50 who received a tail on the flip of the coin had paid a bribe in the past year. Therefore an estimate of the proportion of the sample who had paid a bribe would be 14/50 or 28% [44]. Thus it is possible to estimate the prevalence of a sensitive trait with the RR technique, even though one is not able to find out which of the hand raisers are amongst the bribers.

To illustrate this more formally, consider an interviewer asking an interviewee two statements that are logical opposites[45]:

Statement 1: "I have paid a bribe to a public official in the past 12 months"

Statement 2: "I have not paid a bribe to a public official in the past 12 months"

The interviewee is given a fair randomizing device (e.g. a die) that follows a bernoulli distribution[46] and is instructed to use the device (e.g. by rolling the die) without letting the interviewer see the outcome, and to confirm or deny one of the

---

[44] With 95% confidence intervals: {27.997:28.003} calculated using the variance formula in section 4.2

[45] Part of the following analysis is borrowed from [Fox & Tracy , 1986]

[46] A discrete distribution with 2 possible outcomes: a success (1) occurring with probability p, and a failure (0) with probability $1 - p$

two statements depending on the outcome of the device. For example, in the event of a success (with probability p), respond to statement 1; in the event of a failure (with probability $1 - p$) respond to statement 2.

Using this set-up, the total number of positive (yes) responses (λ) can be described in terms of the probability of being guilty of the sensitive act (π) as follows:

$$\begin{aligned} P(yes) \;=\; & (P(Statement1) * P(yes|Statement1)) + \\ & P(Statement2) * P(yes|Statement2) \end{aligned} \tag{4.2.0.31}$$

Substituting $p \equiv P(Statement1)$ and $\pi \equiv P(yes|Statement1)$ into the above gives:

$$\lambda = p\pi + (1 - p)(1 - \pi) \tag{4.2.0.32}$$

Since p and λ are known to the interviewer, this equation can be solved to find an estimate of the proportion of the sample with the sensitive trait[47] $\hat{\pi}$:

$$\hat{\pi} = \frac{\lambda + p - 1}{2p - 1} \quad for \quad p \neq 0.5 \tag{4.2.0.33}$$

Randomising devices for this method can include a die (if it lands on 1 or 2 then answer question 1; if 3,4,5 or 6 then answer question two, or some other combination such that $p \neq 0.5$).

The RR technique outlined above (and described formally in equations (4.2.0.31) to (4.2.0.33)) was the original method posed by Warner [1965]. Since the initial paper was published many variants of the technique have been developed in order to increase the interviewees' level of truth telling and improve the efficiency of the RR estimator. These different methods include: the unrelated question technique and the forced-response technique. The current study uses the forced-response technique.

The Forced Response Technique The formal model of the forced response technique is statistically equivalent to the unrelated question RR with known population prevalence (see section A.2.1). The interviewee uses a randomising device but this time he/she is instructed to answer the sensitive question truthfully with probability p; to answer "Yes" with probability θ; and "no" with probability $1 - p - \theta$. In this case, the probability of observing a "yes" response is given by:

$$\lambda = p(question1) * p(yes|question1) + p(forced) * p(yes|forced) \tag{4.2.0.34}$$

[47]With sample variance $Var(\hat{\pi}) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2}$ [Lensvelt-Mulders et al. , 2005a]

The estimate for the prevalence of the sensitive trait is given by[48]:

$$\hat{\pi} = \frac{\hat{\lambda} - ((1-p)\theta)}{p} \tag{4.2.0.35}$$

Use of the forced forced response technique could involve a coin: if heads then answer "yes" regardless of the true answer and if tails then answer truthfully. This reduces the need for a second sample[49] and also does not require a second question to be asked. Lensvelt-Mulders et al. [2005a] assess the RR designs mentioned above (Warner's original method; the two unrelated question designs; and the forced response technique) as well as two others and find the unrelated question technique with known population prevalence and the forced response technique to be the most efficient RR designs. The current chapter uses the latter method to compare with the indirect questioning method.

Evaluating The Usefulness Of The Randomised Response Technique In Applied Work On Sensitive Topics The Randomised response technique has been used to estimate the prevalence of illegal drug use [Kerkvliet , 1994b], cheating in classrooms [Kerkvliet , 1994a], corruption [Azfar & Murrell , 2009, Clausen, Kraay & Murrell , 2010, Jensen & Rahman , 2011] and illegal poaching [St. John; Keane; Edward-Jones; Jones; Yarnell and Jones , 2011]. Previous work has been done which compares the RR technique to other survey methods including telephone interviews, self reported questionnaires and direct face-to-face interviews. These studies work on the assumption that a higher rate of admitting to a sensitive act is a more accurate measure than a lower one. The general consensus from this branch of research seems to be that the RR technique is at least as good at getting honest responses as the other methods

Beyond its initial aim of estimating the proportion of a sample with a sensitive trait, the RR technique has also been used in regression analysis to identify the determinants of the sensitive trait [Kerkvliet , 1994a, St. John; Keane; Edward-Jones; Jones; Yarnell and Jones , 2011]. The procedure has also been used as a screening mechanism to identify dishonest interviewees. This is done by using a series of different RR questions and removing the people who respond "No" to every question due to the low probability of this occurring.

Despite its use as a truth revealing mechanism the RR technique suffers from many problems. Many of these problems arise as a result of the randomising process: i.e. attaching the interviewee's response to a probability distribution. Firstly, the introduction of randomness in order to provide anonymity to the interviewee

[48] This estimator has a sample variance of: $\frac{\hat{\lambda}(1-\hat{\lambda})}{np^2}$.

[49] In the unrelated question technique with unknown population variance, two separate samples are required.

leads to an inflated variance for the estimate of the sensitive trait (Var($\hat{\pi}$)). Therefore, the efficiency of the estimator is reduced. Furthermore, there is evidence to show that despite the protection that the randomisation process gives to the interviewee, people still choose to lie about their status in relation to the sensitive question [Azfar & Murrell , 2009]. The problem of dishonesty with the RR procedure means that the estimated proportion of people engaging in the sensitive act (π) might sometimes lie below 0; and conversely the estimated proportion of people *not* engaging in the sensitive act can lie above 1 [Azfar & Murrell , 2009]. Thus, the combination of randomisation and dishonesty can potentially make any RR point estimate meaningless[50]. To be sure, dishonesty also has potential detrimental effects on estimates made from direct questions. However, the mismeasurement that arises from misreported direct questions does not necessarily generate a statistically impossible estimate. Any mismeasured direct questioning (henceforth, DQ) estimate of the proportion of guilty interviewees will lie between 0 and 1 (inclusive).

Another issue with the RR procedure is that, by design, it does not allow anybody to identify those who are admitting to engaging in the sensitive behaviour. With a direct questioning method one can tell who bribes and who (allegedly) does not; however, the RR process hides this by adding random noise to the data. Furthermore, the randomisation process requires an increased sample size in order to be confident in the results. This can be seen from the variance of the estimators:

$$Var(\hat{\pi}_{warner}) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2} \tag{4.2.0.36}$$

$$Var(\hat{\pi}_{forced \quad response}) = \frac{\hat{\lambda}(1-\hat{\lambda})}{np^2} \tag{4.2.0.37}$$

All 3 methods need a relatively large sample size in order to improve the efficiency of the estimator. Many of these problems could potentially disappear if one increased the value of p. That is, if one increased the probability that the interviewee was posed with the sensitive question. However, doing this brings the RR procedure closer to DQ (for DQ: $p = 1$) and reduces the incentive for the interviewee to respond honestly. There seems to be a trade-off between efficiency and information (cooperation) provided by the respondent [Warner , 1965]. Many of the RR designs used in the economic literature seem to ignore some of the general guidelines for asking sensitive questions. These are discussed at the beginning of section 4.2 and include: adding more detail to the question; and increasing the length of the question. Failing to take these strategies into account can also reduce

[50] An estimate of -30% of people paying bribes is not necessarily useful for determining the prevalence of bribery.

the cooperation of the interviewee and distort the estimate of the sensitive trait. Also, despite its use in estimating the determinants of sensitive behaviour, the randomisation process and potential dishonesty provide added noise to the estimates, if dishonesty is correlated with the independent variables, then the coefficients in such a regression will be biased. To the author's knowledge there is no study on whether a randomised response maximum likelihood estimation is superior to a standard maximum likelihood estimation using a direct questioning approach for the dependent variable. However, to the extent that it adds noise to an already (potentially) mismeasured variable, the RR maximum likelihood estimation seems to have more problems to overcome than an ordinary probit or logit model applied to direct questions.

The Angels Assumption And Expected Distributions Instead of using a single RR question to estimate the prevalence of a sensitive trait, some studies have used a series of RR questions, each asking about whether or not the interviewee had a sensitive trait or was guilty of committing a socially undesirable act [Lensvelt-Mulders et al. , 2005a]. A common procedure is to hand the interviewee a fair coin [51] and to give them the following instructions:

> "Please toss the coin handed to you by the enumerator before each question is posed without letting him/her see the results. Always answer YES if the coin comes up HEADS. Answer the question TRUTHFULLY if the coin comes up TAILS (i.e. answer YES if you have done the behaviour; Answer NO if you have never done this behaviour" (ES, Manufacturing Questionnaire. 2007)

The result of this exercise will be a series of positive (yes) and negative (no) responses. Cruyff, Van den Hout & Van der Heijden [2008] label the sum of positive responses the "sum score" of the RR questions. Azfar & Murrell [2009] introduce the concept of the "angels assumption" to investigate the distribution of yes/no responses to a series of independent randomized response questions[52]. The angels assumption is the expected distribution that would arise if no interviewee has the sensitive trait in question and all respondents give honest answers[53]. If there are seven independent randomised responses questions, then the Angels Assumption would be identical to the binomial probability distribution with $p = 0.5$ and $n = 7$. This is shown in Figure 4. Under these conditions, the probability of observing a

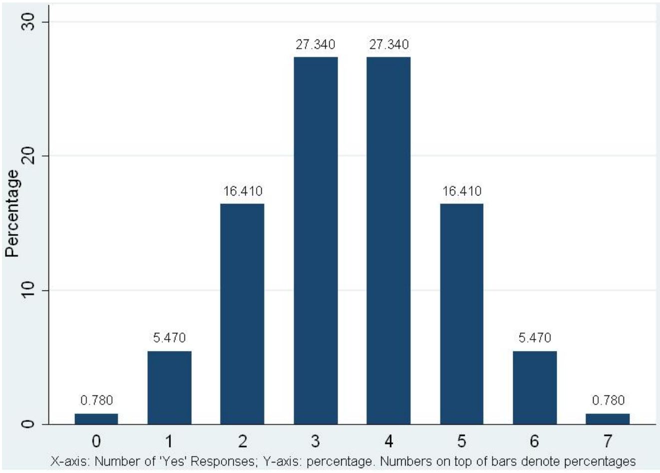
[51] A coin with equal probability of landing on heads and tails: $P(Heads) = P(Tails) = 0.5$

[52] The answer or refusal to answer one question did not have any bearing on whether the interviewee was posed another question; all questions were posed to all interviewees; and all answers were coded as responded by the interviewees.

[53] So no false responses and no non-response

'Yes' is 0.5, $P(Y_i = 1) = 0.5$. Under the Angels Assumption, both the expected number of yeses and nos is 3.5. [54]

Figure 4: Distribution Of 'Yes' Responses For The Angels Assumption



In reality, the responses to a set of RR questions show a positively skewed distribution with much of the sample answering "no" six or seven times. The relatively high number of firms who report 'No' seven times is used as evidence of missreporting. Previous studies have chosen to label the firms who report 'No' seven[55] times (or equivalently; those with a sum-score of 0) as 'reticent' and the rest of the firms as 'Possibly Candid'. Although this method is very likely to identify some of the misreporting, it is possible that some firms labelled as 'Possibly Candid' are in fact reticent. Also, the possibility still exists (though it is relatively small) that some firms labelled as reticent did in fact follow the rules and report honestly.

The current investigation adds to the measure of reticence by seeing how the estimated proportion of reticent respondents change as the threshold for reticence changes. This is done by labelling everybody who reports 6 'Nos' or more as reticent; then 5 'Nos' or more; then 4 'Nos' or more.

In Azfar & Murrell [2009], interviewees were asked ten questions; 7 of which were sensitive and 3 of which were labelled as sensitive but to a lesser degree. The survey included the 3 less sensitive questions in order to allow the firms to say "yes" a few times while still answering "no" to the 7 more sensitive questions. In identifying the firms who misreported their status, previous studies have only used the 7 sensitive questions and have ignored the three less sensitive questions. However, it might be useful to bring the 3 less sensitive questions into the analysis.

[54] The mean, median, and mode for this distribution is 3.5

[55] There are seven sensitive questions in total. So 7 is the maximum number of 'Nos' that a firm can give

Using these 3 extra questions can allow different levels of reticence to be identified. Firms that responded 'No' to all seven sensitive questions *and* the three less sensitive questions might be defined as "Extremely reticent", whereas firms who answered "No" to the seven sensitive questions and "Yes" to the three less sensitive questions might simply be labelled as "Reticent"[56]. This will account for at least two things: 1) The fact that answering "no" seven times in the main questions is nearly impossible under the null-hypothesis of the Angels Assumption; and 2) That, all other things being equal, the probability of answering 'no' 10/10 times (to the 10 sensitive questions) is smaller than the probability of answering 'no' 7/7 times.

Previous literature uses the Angels Assumption as a starting point. The largest difference between the observed responses and that of the Angels Assumption is used as a lower bound estimate for the level of misreporting [Azfar & Murrell , 2009, Clausen, Kraay & Murrell , 2010, Clarke , 2012]. I.e., if, for 3 RR questions, the percentage of "yes" responses are: 35%, 40% and 45%, then the lower bound estimate for reticence is 30% (=2*(50-35)%), so at least 30% if the sample are believed to be reticent [57]

This discussion suggests that whilst the RR provides protection for the interviewee, it is not clear that one can be more confident in the results of a RR estimator compared to estimators from other questioning strategies. The RR suffers from random noise due to the randomising device, an inflated variance, issues with sample size as well as concerns about the accuracy of the estimator. From the literature, it is unclear whether (in light of these problems) it is strictly preferred to other questioning methods. However, there might still be some avenues for further development of the RR technique in the economic literature. Employing some strategies for asking sensitive questions might improve the accuracy of estimates from the RR procedure.

Indirect Questioning Another approach to asking sensitive questions that has been used in the literature is to ask the interviewees indirectly. The indirect questioning approach can take several forms, from asking what somebody like the interviewee would do [Svensson , 2003], to asking what they would do in a hypothetical situation [Kantor , 2003]. The type of indirect question asked can depend on the purpose of the question. Sometimes the aim is to avoid implicating the person of wrongdoing [Sood, Burger, Yoong, Kopf & Spreng , 2011], whilst in other cases the aim is to try to find out the person's true status rather than their

[56] It is important, however, to remember that some of the firms labelled as "Reticent" might indeed be innocent of the acts in question (both sensitive acts and less sensitive acts).

[57] For any individual question, the expected percentage of "yes" responses under the Angels Assumption is 50%.

imagined status

Nega, Mathijs, Deckers & Tollens [2009] use the level of trust to analyse the relationships between gender, social capital and empowerment among households in Ethiopia. Due to trust being potentially difficult to measure, in part because it can be interpreted differently by different people, the authors choose an indirect approach when trying to quantify the level of trust amongst Ethiopian households. An example of a direct questioning approach in the measurement of trust would be to ask: "Generally speaking would you say that most people can be trusted or that you cannot be too careful in dealing with people?". The authors ask this question in the survey but state a preference for a more indirect approach. The interviewers state that "In this village/neighborhood, one has to be alert or someone is likely to take advantage of you" and give the households a choice of five options: agree strongly; agree somewhat; neither agree nor disagree; disagree somewhat; and disagree strongly. In this case, the indirect question approach is used in order to minimise the level of bias that can arise from the households' different interpretations of "trust".

In a study on how conflict and violence affects individuals, households and communities, Brück, Justino, Verwimp & Avdeenko [2010] use indirect questions as a way of protecting the interviewees from the harm associated with recalling a traumatic event. Instead of asking directly if any household member was injured during the previous conflict, the authors suggest asking about individual household members experiences during different time periods - with one time period corresponding to the time of the conflict. By reducing the level of threat to the interviewee the authors hope to reduce the rate of non-response and improve the validity of the data. A potential flaw in this is that interviewees, in an attempt to not relive the traumatic experiences, might give an answer that satisfies the question but is unrelated to the traumatic event. Whilst providing the interviewer with an honest answer, this might fail to answer the question that the interviewer was actually interested in, any subsequent analysis of such questions might wrongly infer the absence of an injury within the household.

Another approach is to ask the interviewee about the behaviour of people "similar to themselves", people like them, or typical people. In the case of company surveys, interviewers might ask about the behaviour of "a firm in your line of business". Such indirect approaches have been used to ask questions about firm-level bribery [Svensson , 2003]; the use of informal labour [Almeida & Carneiro , 2005]; and failing quality standards in companies [Sood, Burger, Yoong, Kopf & Spreng , 2011]. The idea is that by asking what somebody similar to them would do, the question is removing the threat from potentially admitting to a sensitive or illegal act because the interviewee can plausibly claim that they were not referring to their own behaviour but the behaviour of somebody else. Such measures tend to receive a higher response rate and a higher rate of positive responses than direct questions on sensitive questions. One potential flaw of this approach is that it might be difficult to actually identify who the interviewee was referring to, themselves or their peers - this lack of knowledge might add some noise to the response. This can have implications for empirical research, for example if trying to identify the effects of bribery on firm-level growth. Paying a bribe to get a government contract might be beneficial for the firm that pays but detrimental to its competitors, likewise if a “similar firm” pays a bribe, it might be good for that firm but bad for other firms. Thus, using a variable that describes “firms like this one”, where it is not necessarily clear which firm is being referred to, might make empirical analysis of bribery and growth more difficult to conduct.

It might help to distinguish between the mechanisms by which the RR technique and the indirect questioning approach seek to reduce the threat of answering a question and encourage truthful responses. With the indirect approach, the interviewee can plausibly claim innocence and state that they are describing somebody else. With the RR technique, the interviewee can claim innocence and state that they are responding to a different question or instruction. The implication of this is that with the indirect questioning approach, it is not necessarily easy to determine who the interviewee is referring to: themselves or people similar to them. With the RR question it is impossible to determine which question each person was answering, the sensitive one or the non-sensitive one. Therefore, both RR technique and indirect questioning have a degree of uncertainty attached to them.

4.3 Data & Research Questions

This section describes the data used in this study and the research questions that this study addresses. Information used in this section includes RR questions on business operations; indirect questions on company behaviour; opinion based questions on the impact of corruption on business operations; and company characteristics.

4.3.1 Survey Data

Introduction Part of the data used in this study consists of information from a business survey that was carried out as part of a study on the business environment in Nigeria. Firms were asked about different things ranging from their company characteristics to their subjective opinions on the business environments. Firms were also asked a set of RR questions about their company practices as well as indirect questions about company behaviour [World Bank , 2007].

The scope and coverage; sample frame; and selection procedure for the data are as described in Section 2.4.

Randomised Response Questions The specific RR questions used in the survey are outlined in Section 2.4.5. The RR questions in the survey were asked using a variant of the forced response technique described in section 4.2.

Indirect Questions The indirect questions used in the questionnaire attempt to get information from the interviewees by: adjusting the level of detail required in the answers; using longer questions; explaining that the behaviour in question is common; being introduced by a set of warm-up questions; being asked towards the middle/end of the questionnaire, by which time the questioner presumably had established credibility and trust with the interviewee; being placed amongst a series of questions about time spent dealing with government regulations, potentially being less sensitive and more appropriate in such a location; being phrased indirectly in order to avoid implicating the interviewee of any wrongdoing; and being carried out by a non public-sector organisation, potentially increasing participation and attracting the confidence and trust of the participating firms. Details of the specific questions used; and the reasons for their use, are outlined in section 2.4.6.

4.4 Methodology

This section explains the models used in the subsequent analysis. This is the first study to formally model the RR technique for a series of RR questions with reticence. The model is created by considering the likelihoods of responding "Yes" and "No" to a RR question, respectively. After considering these likelihoods, the likelihoods for observing x "yes" responses to a series of RR questions is modelled.

A distinction should be made between the results from a single RR question and a set of RR questions. A further distinction should be made between the results from questioning an individual and the results from questioning a sample of individuals. It is possible to analyse results from a single RR question asked to an individual; a single RR question asked to a sample of individuals; a set of RR questions asked to an individual; and a set of RR questions asked to a sample of individuals.

Table 4.1: Potential Information Achievable With Different Randomised Response Situations

Number Of Questions	Number of Interviewees	Potential Information Achievable
1 Question	1 Individual	Observed response (Y_i)
1 Question	Sample of N > 1 individuals	Estimated average prevalence of sensitive attribute (b)
10 Questions	1 Individual	Reticence of individual (r_i)
10 Questions	Sample of N > 1 individuals	Average reticence level; Average prevalence of sensitive attributes (Within-individual and between individual) (r; $\frac{\sum_{i=1}^{10} b_i}{10}$; $\frac{\sum_{i=1}^{10} b}{10}$)

Analysing the responses of a sample of individuals to 1 RR question will allow one to estimate an interval of the proportion of reticent respondents. Analysing the responses of a sample of individuals to a set of RR questions will allow one to identify the reticent interviewees. The response from a single RR question to an individual does not necessarily provide any useful information to the interviewer. However, the set of responses from a series of RR questions can provide information concerning reticence.

4.4.1 Randomised Response Design & The Estimation Of False Reporting

The RR method used in this study is a variant of the "forced response technique"[58]. Using this method, a respondent is given a fair coin and asked to privately toss the coin[59] before answering each RR question[60]. The interviewee is then instructed to answer "yes" if the result of the coin toss is a heads, and to answer truthfully if the result of the coin toss is a tails. Representing the probability of getting a heads as p; the probability that the company is guilty of the act in question as b; the probability that the person is reticent as r and the observed responses as:

$$Y_i \begin{cases} 1 \text{ if the ith person responds "Yes"} \\ 0 \text{ if the ith person responds "No"} \end{cases} \tag{4.4.1.1}$$

then, in the absence of missclassification error, the probabilities of observing "yes" and "no" responses are:

$$p(YES) = p(Heads) + (p(Tails) \cdot p(guilty)) \tag{4.4.1.2}$$

$$p(NO) = p(Tails) \cdot p(innocent) \tag{4.4.1.3}$$

Which can be expressed as:

$$YES : p(Y_i = 1) = p + (1 - p)b \tag{4.4.1.4}$$

$$NO : p(Y_i = 0) = (1 - p)(1 - b) \tag{4.4.1.5}$$

However, in the presence of misclassification error (reticence), one must control for 1) the people who flip a heads but are dishonest; and 2) those who flip a tails and are guilty but are dishonest. Doing this gives:

[58]See Lensvelt-Mulders et al. [2005b] for a meta-analysis of RR research.

[59]The interviewee is supposed to flip the coin without letting the interviewer see the result of the coin-toss.

[60]The RR questions are dichotomous response questions.

$$YES : p(Y_i = 1) = p + (1-p)b - rp - r(1-p)b \tag{4.4.1.6}$$

$$NO : p(Y_i = 0) = (1-p)(1-b) + rp + r(1-p)b \tag{4.4.1.7}$$

or:

$$YES : p(Y_i = 1) = p + (1-p)b - rp - rb + rpb \tag{4.4.1.8}$$

$$NO : p(Y_i = 0) = (1-p)(1-b) + rp + rb - rpb \tag{4.4.1.9}$$

Re-arranging this gives:

$$p(Y_i = 1) = p(1-r) + (1-p)b(1-r) \tag{4.4.1.10}$$

$$p(Y_i = 0) = (1-p)(1-b) + (p + b - bp)r \tag{4.4.1.11}$$

Where $r = Pr(y^{obs} = 0|y^{true} = 1)$. So, a "yes" is observed if: (the person flips a heads or (the person flips a tails and is guilty)) AND the person is not dishonest. A "no" is observed if: (the person flips a tails and is innocent) or ((the person flips a heads and is dishonest) or (the person flips a tails and is guilty but is dishonest)).

This assumes that firms that are not guilty of the act do not lie and say they performed the act: $q = Pr(y^{obs} = 1|y^{true} = 0) = 0$[61]. This assumption has been maintained in the literature [Jensen & Rahman , 2011]. A justification for this assumption is provided in section 4.4.2.

There is a question of whether there are different types of reticence. I.e. whether the reticence, r, of somebody that is guilty of an act but will still say "no" regardless of the outcome of the coin-toss (br); is different from the reticence r of somebody who is innocent of the act but is unwilling to say "yes" if they get a heads ($pr(1-b)$). In other words, one might want to treat the innocent pleading their innocence as different from the guilty avoiding their guilt. Allowing for different types of r might make a difference in the results. In order to deal with this, it might be useful to recall the literature's definition of a reticent interviewee: as somebody "who gives knowingly false answers with a nonzero probability when honest answers to a specific set of survey questions could lead to the inference that the respondent might have committed a sensitive act" [Azfar & Murrell , 2009]. This definition encompasses both those who are guilty and those who are innocent of the act in question. Therefore, it seems plausible to continue with the analysis using an aggregation of the different possible types of reticence, r.

[61] Another assumption is that the probability of being reticent is independent of both the probability of committing the act; and the result of the coin-toss. So: $r \| b, p, (1-p)$.

Furthermore, when reticence is considered as misclassification error (in the framework of Hausman, Abrevaya, Scott-Morton [1998]), it seems reasonable to use one single reticence measure [62].

If the probability that a particular company is reticent, r_i, is the same for all firms; and that this is equal to the proportion of firms in the sample that are labelled as reticent, r; so $r_i = r \quad \forall \quad i$. Then an estimate of b can be obtained by:

$$b = \frac{\lambda - p(1-r)}{(1-p)(1-r)} \tag{4.4.1.12}$$

where λ represents the proportion of people who respond "Yes"[63]. This gives the average level of guilt (bribery/tax evasion) for all firms.

This study adds to the literature by introducing validation data for b and estimating the level and behaviour of the misclassification error, r, using the formula:

$$r = max\left\{0, 1 - \frac{\lambda}{p + (1-p)b}\right\} \tag{4.4.1.13}$$

The current study uses the forced-response technique as described in section 4.2. Since the procedure uses a fair coin, $p = 0.5$ and Equation 4.4.1.13 becomes:

$$r = max\left\{0, 1 - \frac{2\lambda}{1+b}\right\} \tag{4.4.1.14}$$

4.4.2 Justification For Treating Probability Of False-Positives As Zero

This chapter treats the probability of observing a "1" when the true value is a "0" as zero (i.e. that $q = 0$). This is the assumption that is maintained in the literature [Artís, Ayuso & Guillén , 2002]; and there is both intuitive and empirical support for working with this assumption, which is explained below.

The intuitive justification for assuming that the error from miscoding a 0 as a 1 is zero is because people will not confess to anti-social acts if they did not do it. This result occurs even with the use of the RR technique [Lensvelt-Mulders et al. , 2005b]. This is due to the psychic cost of admitting to an illegal act as well as the small probability of leakage.

The empirical support for setting $q = 0$ is observed in the literature [Azfar & Murrell , 2009]. When observing responses from a set of binary choice questions, the distribution of responses tails out to the right[64] and the proportion of observed positive responses is lower than the expected value of positive responses. This lends

[62] Certainly, misclassification can arise from different sources.

[63] The sample variance of this estimator when there is no reticence, ($r = 0$), is: $Var(\hat{b}) = \frac{\hat{\lambda}(1-\hat{\lambda})}{N} \cdot \frac{1}{p^2}$ ([Lensvelt-Mulders et al. , 2005b]).

[64] This implies that $r > q$.

support to the idea that there is not much force pushing people to response "Yes" to these questions. Taken as a whole, these factors give credibility to setting $q = 0$ and continuing with an analysis of the effects of r.

4.5 Results

4.5.1 The Prevalence Of Sensitive Behaviour

Table 4.2 shows the 10 RR questions used in the survey and the percentage of interviewees who responded yes/no to these questions. The values are then transformed into estimates of the prevalence of these acts using the equation: $b = max\{0, \frac{\lambda - 0.5}{0.5}\}$, where b is the proportion of guilty people and λ is the proportion of "Yes" responses to the respective RR questions. This gives the unadjusted[65] estimate for the prevalence of these behaviours. If the estimate of the prevalence is negative, a zero has been placed in the appropriate cell, this occurs on seven occasions (Questions: 2; 3; 5; 6; 8; 9; and 10). A negative estimate for the prevalence of an act can be a red flag for the presence of misreporting amongst the respondents. If there were no misreporting the expected domain for the values of λ would be between 0.5 and 1. Any value of λ statistically less than 0.5 is a potential sign of underreporting of the act in question. The highest number of positive ("Yes") responses were generated by question 7, the question concerning lying in one's self interest. 51.8% of firms admitted to doing this, suggesting an unadjusted estimate of 3.6% of firm managers who lie in their self interest.

Focusing on the last row of Table 4.2, 36.1% of people answered "Yes" when asked "Have you ever unfairly dismissed an employee for personal reasons?". This suggests that at least an estimated 13.9%[69] of the interviewees responded "no" when they flipped a heads when they should have said "yes". Since half the sample would have tossed a heads, this implies an estimate of at least 27.8% reticent respondents. This is how the final column in Table 4.2 is calculated. One must note that this estimate is a lower bound estimate for the level of reticence because it assumes that nobody is guilty of the act (nobody fired an employee for personal reasons). If 10% of the firms were actually guilty of firing an employee for personal reasons, then (using the question $r = max\{0, 1 - \frac{2\lambda}{1+b}\}$) the estimate of

[65] This estimate is labelled as unadjusted because it does not take misreporting into account.

[64] Confidence Intervals are calculated using the variance for the RR estimator as shown in Lensvelt-Mulders et al. [2005b].

[65] This lower bound is based on the assumption that every interviewee is innocent of the act in question.

[66] This upper bound is based on the assumption that every interviewee is guilty of the act in question.

[69] (50-36.1)%

Table 4.2: Observed Responses and Estimated Percentage Of Transgressions

	Question	Observations (%)		Estimates (%)		95% Confidence Intervals[66]		Lower Bound Estimate Of Reticence(%)[67]	Upper Bound Estimate Of Reticence(%)[68]
		"Yes"	"No"	"Yes"	"No"				
1	**Have you ever paid less in personal taxes than you should have under the law?**	**50.3**	**49.7**	**0.6**	**99.4**	**0.5994**	**0.6006**	0	49.7
2	**Have you ever paid less in business taxes than you should have under the law?**	**42.5*****	**57.5**	**0^C**	**100.0^C**	**0^C**	**0.0006**	15	57.5
3	**Have you ever made a misstatement on a job application?**	**42.9*****	**57.1**	**0^C**	**100.0^C**	**0^C**	**0.0006**	14.2	57.1
4	Have you ever used the office phone for personal businesses?	50.5	49.5	1.0	99.0	0.9994	1.0006	0	49.5
5	**Have you ever inappropriately promoted an employee for personal reasons?**	**40.4*****	**59.6**	**0^C**	**100.0^C**	**0^C**	**0.0006**	19.2	59.6
6	**Have you ever deliberately not given your suppliers or clients what was due them?**	**37.9*****	**62.1**	**0^C**	**100.0^C**	**0^C**	**0.0006**	24.2	62.1
7	Have you ever lied in your self interest?	51.8^{++}	48.2	3.6	96.4	3.5994	3.6006	0	48.2
8	**Have you ever inappropriately hired a staff member for personal reasons?**	**40.7*****	**59.3**	**0^C**	**100.0^C**	**0^C**	**0.0006**	18.6	59.3
9	Have you ever been purposely late for work?	47.8***	52.2	0^C	100.0^C	0^C	0.0006	4.4	52.2
10	**Have you ever unfairly dismissed an employee for personal reasons?**	**36.1*****	**63.9**	**0^C**	**100.0^C**	**0^C**	**0.0005**	27.8	63.9

Number of Observations: 3,200. Questions in bold highlight the relatively more sensitive questions (Clausen et al., 2010). *($^+$) stars (crosses) display the significance of one-sided binomial test that the observed frequency is smaller (larger) than 50.0% (*($^+$)=10% level, **($^{++}$)=5% level, ***($^{+++}$)=1% level). P-value=0.000 for questions 2,3,5,6,8,& 10. P-value=0.008 for question 9. P-value=0.019 for question 7. ^{C} denotes a censored figure.

reticent respondents is 34.4%. If, as per Clausen, Kraay & Murrell [2010][70], one treats firm managers who are reticent in answering one question as being reticent on all other questions as well then the results from this table give an estimated lower bound for the percentage of reticent respondents of 27.8% [71]. These results are illustrated in Figures 5 & 7 which show the response to each question and the sum score of the sensitive questions respectively.

Table 4.3 shows the results from the indirect questions that allow for a categorical response. In each case, the categories are: strongly agree; tend to agree; strongly disagree; tend to disagree. The majority of firms did not believe that bribery was common. However, the majority of those who did agree that it was prevalent also said that firms know in advance how much payment is expected from them. This tends to agree with the notion of corruption being prevalent within

[70] I.e. setting $\underline{r} = max\{0, 1-2\lambda_1, 1-2\lambda_2, ..., 1-2\lambda_{10}\}$, where subscripts refer to the question number and $\underline{r}$ is the lower bound estimate for reticence.

[71] I.e. the highest lower bound estimate.

the business environment in Nigeria [Smith , 2007].

Table 4.3: Results From Indirect Questions With Categorical Responses

	To what extent do you agree or disagree with the following statements?		
	It is common for establishments in this line of business to have to pay informal payments/gifts to get things done with regard to customs, taxes, licenses, regulations, etc.	Establishments in this line of business know in advance about how much this informal payment/gift is to get things done.	(Firms who answered "strongly agree/tend to agree" to question 1) Establishments in this line of business know in advance about how much this informal payment/gift is to get things done.
Strongly agree	8%	7%	13%
Tend to agree	30%	27%	54%
Tend to disagree	38%	38%	21%
Strongly disagree	24%	28%	12%
N	3199	3198	1215

Results from the rest of the indirect questions are shown in Table 4.4. In this final set of questions firms were asked to report bribes, hidden sales, and hidden workforce[72]. They were allowed to report either in percentage of sales or Naira (for bribes) and percentage of total amount (for hidden sales and hidden workforce). The data in Table 4.4 is generated by creating a dummy variable equal to 1 if the firm reported a positive amount of bribery or hidden activity and 0 otherwise. With the indirect questions on bribery, firms were asked: i) what percentage of total annual sales, or estimated value, that similar establishments pay in informal payments/gifts to public officials to speed regulatory processes or acquire government services; and ii) what percentage of contract value would be paid in informal payments/gifts in order to secure a government contract. For these two questions, the dummy variable takes a value of 1 if the firm reports a positive value, and 0 otherwise. For the questions concerning hidden activity, firms were asked: i) what percentage of total annual sales that similar firms in the same sector of activity report for tax purposes; and ii) what percentage of the total workforce similar establishments declare for tax purposes. For these variables, the dummy takes a value of 1 if the firm reported less than 100% (suggesting the presence of hidden activity), and 0 otherwise.

Table 4.4 shows that 71% of firms reported that they did not declare some amount of sales for tax purposes. A similar percentage of firms (68%) said that some of their workforce were not declared for tax purposes. In total, 81% of firms reported that firms like them did not declare at least one of the two activities for tax purposes, and 32% reported not declaring either one or the other (but not both). From these results it seems that indirect questioning generates a higher percentage of honest responses than randomised response questions. Recalling the

[72]Questions are listed in section 4.4.

Table 4.4: Indirect Questions Concerning Bribe Payments And Hidden Sales/Workforce

	Bribe For General Government Services	Bribe For Government Contracts	Unreported Sales	Unreported Workforce
Yes	53%	57%	71%	68%
No	47%	43%	29%	32%

RR questions on business taxes in Table 4.2, the estimated percentage of firms who paid less in business taxes than they were supposed to under the law was 0%. However, by asking questions indirectly, at least 81% of firms have hidden sales or workers from their tax forms. To be sure, the IQ asks about sales/employees reported for tax purposes while the RRQ asks about what is paid in business taxes. For the comparison to be valid it must be the case that when firms reported about the amount of business tax paid/unpaid, they were basing their answer in relation to the amount that should have been paid and not on the amount (of sales or employees) that they had reported for tax purposes.

Nevertheless, another result which favours indirect questioning over randomised response questioning is the relatively large disparity in admitting between the two techniques. Indirect questioning generates a minimum level of admitting of 53% while the maximum estimate of guilt from the randomised response questions is 3.6%. This tends to suggest that people favour the reduced stigma that comes from the longer questions, face-saving wording, and projective nature of the indirect questions compared to the randomisation process of the randomised response questions. One must note that this survey only includes a RR process with a probability of selection, p, of 0.5. It is possible that randomising devices with a lower probability of selecting the sensitive question might yield more candidness on the part of the interviewees. However, such a design would reduce the efficiency of any estimate of the prevalence of the sensitive behaviour in question (see Equation 4.2.0.37) and require a larger sample size to provide useful results.

4.5.2 The Distribution Of Responses

Table 4.5 shows the distribution of responses for Romania[73]; Nigeria; and the Angels Assumption. The Angels Assumption [Azfar & Murrell , 2009] represents the expected distribution of answers if the following conditions hold: Everybody is innocent of the sensitive act in question; there is no non-response; and everybody follows the procedure by answering truthfully. If these conditions are satisfied then the distribution of responses to the seven sensitive questions follow a binomial distribution with p=0.5; n=7. The large proportion of Nigerian & Romanian

[73] Data from Azfar & Murrell [2009].

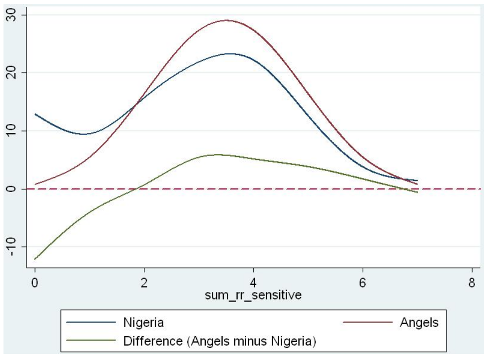
Table 4.5: Distribution of Observed Responses

Frequency of Yes Responses	Binomial Probability Distribution with p=0.5	Angels Assumption	Romania (N=514)	Nigeria (N=3200)
0	0.0078	0.78	10.5	12.9
1	0.0547	5.47	7.8	9.5
2	0.1641	16.41	18.0	15.7
3	0.2734	27.34	20.4	21.9
4	0.2734	27.34	20.9	22.2
5	0.1641	16.41	14.1	12.6
6	0.0547	5.47	7.1	3.8
7	0.0078	0.78	1.8	1.4

Nigerian firms show more reticence (as defined by Azfar & Murrell [2009]) than Romanian ones

firms who never answered "Yes" to any of the questions compared to the Angels Assumption[74] implies the presence of reticence ($r \neq 0$). Furthermore, the higher proportion of firms who answered "No" 6 or 7 times (out of a total of 7 questions) in Nigeria suggests that Nigerian firms are more reticent than Romanian ones.

Figure 5: Comparing The Distributions Of Responses



N=3200. Dashed line==0. The distribution of responses to the seven sensitive questions is a discrete distribution; the outcome of interest can only be an integer value. For convenience, this diagram displays the distribution of responses as continuous.

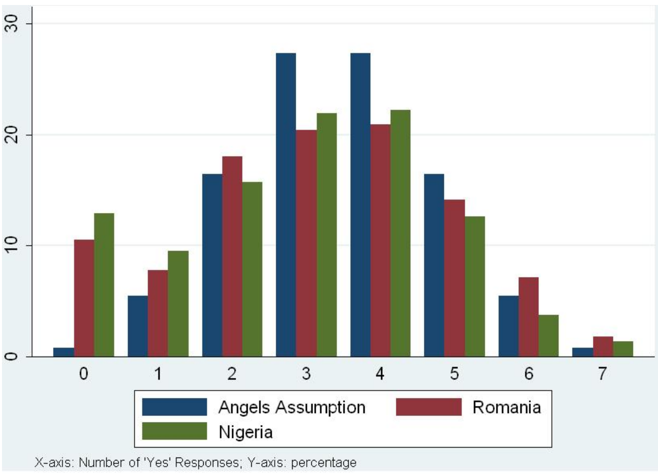
This information is also shown in Figures 5 and 6. Figure 5 compares the expected distribution of responses under the Angels Assumption to the observed distribution of responses from the Nigerian sample; it also shows the difference between the two. Figure 6 shows a bar chart combining Nigeria, Romania, and the Angels Assumption. The x-axis shows the number yes responses (moving from left to right). The y-axis represents the percentage of firms. For both Nigeria and Romania there is initially a relatively high percentage of firms who report 7 "No" responses, after which this percentage drops and the distribution follows a more symmetrical pattern, tailing off to the right. Taken in full, this distribution of

[74] The top row of numbers.

responses represents a departure from the Angels assumption. Since non-response to the RR questions were minimal, this means that there is a high probability that at least some firms were not answering truthfully.

A general result seems to be coming out of the information contained in Figure 5 and Table 4.5. There seem to be two things determining the distribution of responses. Reticence seems to be driving the distribution to the left, creating more "nos" than would be expected under the null hypothesis of the Angels Assumption. Also, guilt seems to be driving some of the distribution to the extreme right. This is because there are more firms who answered "yes" seven times than would be expected under the null hypothesis of the Angels Assumption. So there appears to be at least some firms who committed some of the acts in question and truthfully report this.

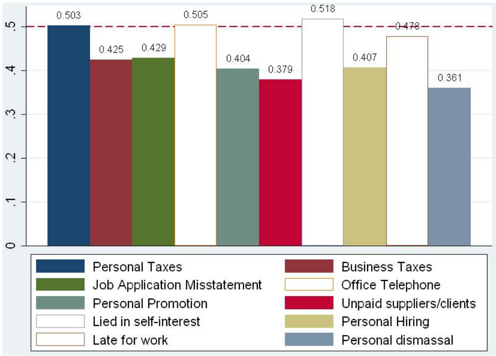
Figure 6: Distribution Of 'Yes' Responses For The Angels Assumption; Romania; And Nigeria



This result is also shown in Figure 7 which shows the observed proportion of "Yes" responses for each question. The dashed line represents the expected proportion of "Yes" responses under the Angels Assumption (50% for each question). Bars with peaks which are significantly less than this value imply the presence of underreporting. The empty (white) bars represent the less sensitive questions and the filled bars represent the relatively more sensitive questions. In general, fewer "yes" responses come from the more sensitive questions.

Due to the presence of misreporting by some companies, it might be useful to temporarily ignore these firms and look at the distribution of responses for the truncated sample. One route to take is to estimate the distribution of responses ignoring those who said 'No' 7 times. Another route is to drop those that said 'No' 7 times and re-estimate the distribution of responses. Both avenues are considered

Figure 7: Proportion Of 'Yes' Responses For Each Question



N=3200. Dashed line represents the expected value under the Angels Assumption.

in Table 4.6 and Figure 8.

Table 4.6: Comparing The Truncated And Reduced Distributions

Number of 'Yeses'	Observed Responses	Binomial Probability Distribution(p=0.5; n=5)	Truncated Sample(No Reticent)	Binomial Probability Distribution(p=0.4, n=6)
0	(0.1294)	(0.0078)	-	0.0467
1	0.0947	0.0547	0.1088	0.1866
2	0.1572	0.1641	0.1805	0.3110
3	0.2188	0.2734	0.2513	0.2765
4	0.2222	0.2734	0.2552	0.1382
5	0.1256	0.1641	0.1443	0.0369
6	0.0384	0.0547	0.0441	0.0041
7	0.0137	0.0078	0.0158	-

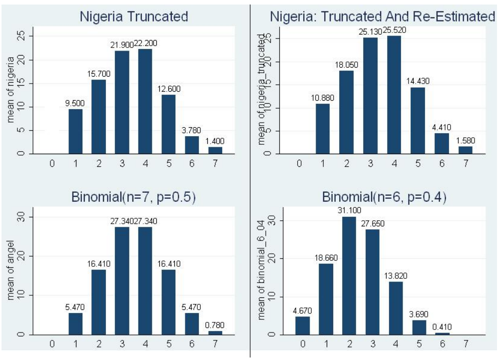
Focusing on the initial set of responses but ignoring those who said 'No' 7 times (the left panel of Figure 8), the distribution of responses appears to be similar to that of a Binomial probability distribution with n=7 and p=0.5 (ignoring the first category). Similarly, if one believes that the reticent are all to be discarded from the data entirely and were to re-estimate the distribution of responses (the right-hand panel) the remaining data appears to be similar to that of a binomial distribution with n=6 and p=0.4 (but shifted to the right)[75].

4.5.3 Deriving Estimates Of Reticence And Guilt From The Likelihood Function

Recalling the estimated probabilities of observing a "1" and a "0" from the randomised response procedure:

[75] This result was obtained by comparing the data with that contained in the table of binomial probabilities.

Figure 8: Comparing Distributions Of Responses



N=3200.

$$p(Y_i = 1) = p(1-r) + (1-p)b(1-r) \tag{4.5.3.1}$$

$$p(Y_i = 0) = (1-p)(1-b) + (p+b-bp)r \tag{4.5.3.2}$$

Table 4.7 shows a set of expected probabilities ($p(Y_i = 1)$) given a set of values for reticence, r, and guilt, b. The table allows r and b to each take on values of 0; 0.5 and 1. Given the 9 different combinations of values for r and b, the expected probabilities of observing a "yes" response fall among the following values: {0; 0.25; 0.375; 0.5; 0.75; 1}. Table 4.8 shows the distribution of responses that would be observed when the probability of observing a "yes" from a single question is 0.75, 0.375, and 0.25 respectively.

Table 4.7: Expected Probabilities Under Different Assumptions

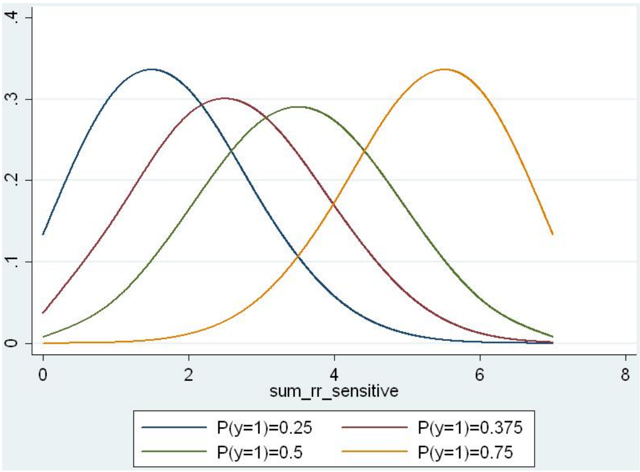
Guilt / Honesty	Everybody is guilty (b=1)	50% are guilty, 50% are innocent (b=0.5)	Everybody is innocent (b=0)
Nobody gives honest responses (r=1)	100% "No" responses $p(Y_i = 1) = 0$	100% "No" Responses $p(Y_i = 1) = 0$	100% "No" responses $p(Y_i = 1) = 0$
50% are reticent, 50% are honest (r=0.5)	$p(Y_i = 1) = 0.5$	$p(Y_i = 1) = 0.375$	$p(Y_i = 1) = 0.25$
Everybody gives honest responses (r=0)	100% "Yes" responses $p(Y_i = 1) = 1$	$p(Y_i = 1) = 0.75$	Angels Assumption (binomial probability distribution with p=0.5and n=7) $p(Y_i = 1) = 0.5$

Recalling the actual distribution of responses that is observed in the data: 0 (12.9%); 1 (9.5%); 2 (15.7%); 3 (21.9%); 4 (22.2%); 5 (12.6%); 6 (3.8%); 7 (1.4%); it is possible to turn the preceding analysis upside down by finding the values of r and b that are most likely to generate the observed data. This is done in the rest of this subsection.

Table 4.8: Distributions of "Yes" Responses Under Different Levels Of Reticence And Guilt

	$p(Y_i = 1) = 0.75$	$p(Y_i = 1) = 0.375$	$p(Y_i = 1) = 0.25$
0	0.01	3.73	13.35
1	0.13	15.64	31.15
2	1.15	28.16	31.15
3	5.77	28.16	17.30
4	17.30	16.90	5.77
5	31.15	6.08	1.15
6	31.15	1.22	0.13
7	13.35	0.10	0.01

Figure 9: Some Distributions For Different Probabilities of Positive Responses



The answer to a single RR question can be modeled as a dichotomous random variable that adopts a value of 1 with probability:

$$p(1-r) + (1-p)b(1-r) \tag{4.5.3.3}$$

and adopts value of 0 with probability:

$$(1-p)(1-b) + (p+b-bp)r \tag{4.5.3.4}$$

The distribution of this random variable is characterised by the parameters: n, the number of trials (the number of RR questions); and p, the probability of a "yes" in a single trial. The mean of this random variable is:

$$n\,(p(1-r) + (1-p)b(1-r)) \tag{4.5.3.5}$$

In the current study $n = 7$. Therefore, the information concerning the probability mass function (pmf) for the answer to the RR question can be described as:

$$f(1) = p(1-r) + (1-p)b(1-r) \tag{4.5.3.6}$$

and

$$f(0) = (1-p)(1-b) + (p+b-bp)r \tag{4.5.3.7}$$

Where $f(\cdot)$ denotes the probability distribution function. With the design of the current investigation it is known that $p = 0.5$, therefore the pmfs become:

$$f(1) = 0.5(1+b)(1-r) \tag{4.5.3.8}$$

and

$$f(0) = 0.5((1-b) + (1+b)r) \tag{4.5.3.9}$$

Given a set of responses to a series of RR questions, e.g.: {1,0,0,1,1,0,1}; the joint probability of observing 4 ones and 3 zeros, ignoring the order of responses, [76] [77] is given by:

$$\begin{aligned} f(1,1,1,1,0,0,0) = \quad & f(1) \times f(1) \times f(1) \times f(1) \times f(0) \times f(0) \times f(0) \\ = \quad & (0.5(1+b)(1-r)) \times (0.5(1+b)(1-r)) \\ & \times (0.5(1+b)(1-r)) \times (0.5(1+b)(1-r)) \\ & \times (0.5((1-b)+(1+b)r)) \times (0.5((1-b)+(1+b)r)) \\ & \times (0.5((1-b)+(1+b)r)) \\ = \quad & (0.5(1+b)(1-r))^4 * (0.5((1-b)+(1+b)r))^3 \end{aligned} \tag{4.5.3.10}$$

It is possible to calculate the value of r that maximises the joint probability of observing {1,0,0,1,1,0,1}. This can be done numerically by choosing different values for r and seeing which one maximises the joint density $f(1,1,1,1,0,0,0)$. It can also be done analytically by interpreting the joint density (Equation (4.5.3.10)) as the likelihood function; taking the derivative of this with respect to r; equating this to zero; and solving for r. The solution for the general case is as follows:

Using the same notation and also defining $n \equiv$ the total number of RR questions asked; and $s \equiv$ the number of questions that a firm answers "yes" (sum-score). The likelihood for a set of RR questions can be described as:

$$L = ((p + (1-p)b)(1-r))^s \times ((1-p)(1-b) + r(p + (1-p)b))^{n-s} \tag{4.5.3.11}$$

The Log-likelihood function is:

[76] 4 "yeses' and 3 "nos".

[77] Assuming that the observations are independent.

$$logL = slog(p+(1-p)b) + slog(1-r) + (n-s)log((1-p)(1-b) + r(p+(1-p)b)) \tag{4.5.3.12}$$

With first order condition for a maximum:

$$\frac{\partial logL}{\partial r} = \frac{-s}{1-r} + \frac{(n-s)(p+(1-p)b)}{(1-p)(1-b)+r(p+(1-p)b)} = 0 \tag{4.5.3.13}$$

$$r = 1 + \frac{s}{n((b(p-1)-p))} \tag{4.5.3.14}$$

Since $p = 0.5$ and $n = 7$ this reduces to:

$$r = 1 - \frac{2s}{7(b+1)} \tag{4.5.3.15}$$

Recalling Figure 5 and using the data from Table 4.5, one can calculate the mean number of yes responses, s. This works out as: 2.91. So, on average, firms said "yes" 2.91 times out of 7. Inserting this value into equation 4.5.3.15 gives:

$$r = 1 - \frac{5.82}{7(b+1)} \tag{4.5.3.16}$$

Using this formula allows one to calculate a maximum likelihood (ML) estimate of the level of reticence required to produce the distribution of responses in Figure 5 & Table 4.5 under different levels of guilt. Some results for this estimation are shown in Table 4.9[78]

Table 4.9: Maximum Likelihood Estimates For Reticence

b	r	$P(Y_i = 1)$
0	16.9%	0.416
0.25	33.5%	0.416
0.5	44.6	0.416
0.75	52.5%	0.416
1	58.4%	0.416

So, under the assumption of innocence from any transgression; the level of reticence would have to be 16.9% to produce the distribution of responses observed in the data.

[78] The second order condition for a maximum is:

$$\frac{-s}{(1-r)^2} - \frac{(7-s)(p+b(1-p))^2}{((1-p)(1-b)+r(p+(1-p)b))^2} < 0 \tag{4.5.3.17}$$

Which is always true for any value of b when $p = 0.5$ and $s = 2.91$.

Alternatively, given a set of values for the level of reticence, one can try to solve the log-likelihood function to get an estimate of the level of guilt. To do this, a similar procedure can be followed, by maximising equation 4.5.3.12 with respect to b instead of r. Doing this gives:

$$\frac{\partial logL}{\partial b} = \frac{(1-p)s}{p+b(1-p)} + \frac{((1-p)r+p-1)(n-s)}{(p+b(1-p))r+(1-b)(1-p)} = 0 \qquad (4.5.3.18)$$

With $s = 2.91$; $n = 7$; and $p = 0.5$ this turns into:

$$b = \frac{-291}{350(r-1)} - 1 \qquad (4.5.3.19)$$

Using this equation, one can derive ML estimates for the level of guilt within the sample. This is done in Table 4.10.

Table 4.10: Maximum Likelihood Estimates For Guilt

r	b	$P(Y_i = 1)$
0	-16.9%	0.416
0.1294	-4.4%	0.416
0.395	37.4%	0.416
0.5	66.3%	0.416

The second value for reticence in Table 4.10 was selected by calculating the proportion of the sample with a sum score of 0/7 (Table 4.5). This figure of 12.8% corresponds to a level of guilt of -4.4%. Since, it is impossible to have a negative level of guilt this results suggests that this estimate of reticence is too low to generate the observed distribution of responses. This same reasoning applies to the first reticence figure of zero.

4.5.4 Predicting The Last Response

Table 4.11: Answer To The 7th Sensitive Question; by Number of previous positive answers

H ave you ever unfairly dismissed an employee for personal reasons?								
Number of Yeses	0	1	2	3	4	5	6	Full Sample
No	93	73	63	56	56	54	30	64
Yes	7	27	37	44	44	46	70	36
N	446	372	641	817	634	227	63	3200

Figures for responses are in percentages.

The previous section showed that there is still a degree of underreporting of behviour even with the perceived protection of the randomised response technique. One reason for this could be that some respondents are unwilling to admit to an undesirable trait no matter what the actual state of affairs. This might be true for some of the firms that responded "No" to every single question, these people might be guilty of the sensitive acts and be trying to deny their guilt; or they might be innocent of the acts and be attempting to proclaim their innocence. Another thing to note is that more people had a sum score of 7 (7 yeses) than would be expected under the null of the angels assumption, so despite the downward bias that occurs when asking sensitive questions, at least some of the sample seem to be following the instructions[79].

The RR questions were asked as a series of questions with binary responses. It is possible that some interviewees decide to tailor their latter responses based on their earlier responses. Managers might be strategic in this manner because they do not want to be discovered as cheating. This study tests this proposition by predicting the response to the final sensitive question based on the the responses to the previous 6 questions. Table 4.11 shows the answers to the final question, disaggregated by the number of positive responses to the first 6 sensitive questions. The table seems to suggest a non-decreasing relationship between the number of previous questions answered in the affirmative, and the probability of answering the last question positively. In other words, firms that responded positively to the first 6 questions also seemed to say "yes" the final sensitive question as well.

The probit estimations in Table 4.12 seem to agree with this result. In all models within this table the dependent variable is the answer to the final RR question. The estimation controls for the firm specific variables that were used in Clausen, Kraay & Murrell [2010] including the gender of the owner; the age group of the owner; the level of formal education achieved by the owner; the sector of the firm; the size of the firm; the region of the country that the firm is located; and the wave in which the survey was conducted. These variables were chosen to estimate the correlates of reticence. In the first model (columns (1) & (2)) the independent variable of interest is the sum score from the previous 6 sensitive RR questions. Results from this estimation show that the response to the final question is highly correlated with the sum score of the previous 6 sensitive questions. The response to the last question is positively correlated with the number of 'yeses' to the previous 6 sensitive questions that the firm answered.[80] The probit coefficient on the sum score variable is 0.226, this is significant at the 1% level. The marginal effect for

[79] Another explanation could be that they were not taking the exercise seriously and simply reported "yes" to every question. However, due to the potential costs of admitting to a tabooed act, this explanation is ruled out (see section 4.4.2).

[80] The correlation coefficient for these two variables is 0.255 and this is significant at the 1% level.

this variable is 0.083. This means that answering one more of the first 6 sensitive questions in the affirmative is associated with an increased probability of answering the last question with a 'yes' of 0.083 probability points (0.082 probability points for the logit model). At the extremes, the difference in probability of answwering the last questiong with a 'yes' is 0.50 probability points higher for somebody who answered all of the 6 sensitive questions in the affirmative compared with somebody who answered 'no' to all of the first 6 sensitive questions.

The second model of Table 4.12 (columns (3) & (4)) separates the sum scores into dummy variables for each group (Sum score={6;5;4;3;2;1;0}). The category "sum score=0" is excluded from this estimation and used as a base category. Results suggest that, relative to those with a sum score of 0, positive sum scores are related to an increased probability of answering "yes" on the final question. Furthermore, the size of the coefficients increase with the size of the sum score. All coefficients are statistically significant at the 1% significance level. Wald tests of the equality of these coefficients reject the null of equal coefficients of the groups with a sum score of 6 and 5 at the 1% level (Prob.>chi2=0.0011). The coefficients on the groups with a sum score of 5 and 4 are not statistically different using this test. The coefficients on groups 4 and 3 are also not statistically distinguishable using this test[81]. Groups 3 and 2 are statistically different at the 10% significance level (Prob.>chi2=0.068); and groups 2 and 1 (Prob.>chi2=0.005) are significantly different at the 1% level. The same results remain when using a logit model to estimate the parameters.

The final model within this table, uses the dichotomous (0/1) responses to the individual questions (questions 1 to 6) as the independent variables of interest. The coefficients on all of the dummies are positive and statistically significant at the 1% level except for that of question 3 (job application misstatement), which is positive but not statistically significant.

This study now turns to the task of comparing predicted outcomes with actual outcomes. In total, 36% of the sample asnwered the last question positively and 64% answered "No" to this question. Setting $\hat{y} = 1$ if $F(x'\beta) > 0.5$ and $\hat{y} = 0$ if $F(x'\beta) \leq 0.5$, one can measure goodness of fit as the percentage of correctly classified observations. This is done in Table 4.13 which compares fitted values and actual values from the first model of Table 4.12. The percentage of correctly specified values in this case is 64.16[82]. In this case, 233 observations are misclassified as 1 when the correct classification is 0, and 878 values are misclassified as 0 when the correct value is 1. The remaining 1762+227=1989 observations are correctly specified.

[81] The coefficients for groups 3 and 5 are also not statistically different from each other

[82] The percentage of correctly specified values from the second and third probit models are 64.77% and 64.45%, respectively.

Table 4.12: Probit Estimations Of Response To Final Question On Results From Previous Questions And Explanatory Variables Used In Clausen at al. (2010)

	(1) Estimation	(2) Marginal & Impact Effects	(3) Estimation	(4) Marginal & Impact Effects	(5) Estimation	(6) Marginal & Impact Effects
Sum Score: Q1-Q6	0.226***	0.083***				
	(0.015)	(0.006)				
Sum Score=1 (d)			0.844***	0.326***		
			(0.115)	(0.043)		
Sum Score=2 (d)			1.149***	0.433***		
			(0.104)	(0.036)		
Sum Score=3 (d)			1.275***	0.474***		
			(0.101)	(0.034)		
Sum Score=4 (d)			1.301***	0.485***		
			(0.104)	(0.034)		
Sum Score=5 (d)			1.353***	0.496***		
			(0.125)	(0.036)		
Sum Score=6 (d)			1.969***	0.612***		
			(0.191)	(0.027)		
Question 1 (d)					0.291***	0.107***
					(0.052)	(0.019)
Question 2 (d)					0.310***	0.115***
					(0.051)	(0.019)
Question 3 (d)					0.036	0.013
					(0.050)	(0.018)
Question 4 (d)					0.290***	0.108***
					(0.050)	(0.019)
Question 5 (d)					0.294***	0.110***
					(0.050)	(0.019)
Question 6 (d)					0.132***	0.049***
					(0.049)	(0.018)
Constant	-0.946***		-1.446***		-0.943***	
	(0.110)		(0.141)		(0.111)	
Other control variables	YES		YES		YES	
Pseudo R-squared	0.057		0.075		0.063	
Observations	3100		3100		3100	
Log-Likelihood	-1904.4		-1868.4		-1891.6	
Chi-Squared	244.0		253.4		264.2	
P Value	0.000		0.000		0.000	

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dependent variable is the response to the last RR question (0=no; 1=yes). (d) for discrete change of dummy variable from 0 to 1. Other control variables include: gender of owner (dummy); age of owner (category); education of owner (category); industry of firm (category); size of firm (dummy); region (dummy); and wave (dummy). The coefficients on these variables are not statistically significant.

Table 4.13: Comparing Actual Outcomes With Predicted Outcomes From Probit & Logit (Model 1) Estimations

	True Outcome		
Predicted	No	Yes	Total
No	1762	878	2640
Yes	233	227	460
Total	1995	1105	3100

Using Table 4.11 to generate predictions with the same method produces similar results. For each sum-score group, if the proportion of people who said "Yes" is less than or equal to 0.5, then the predicted value for each member of that group is set to 0; and conversely, if the proportion of people in a group who said yes is greater than 0.5, then the predicted value for those firms are set to 1. Using this method, all those who said "yes" 6 times are predicted to say "yes" to the final question, while all other firms are predicted to say "No" to the final question. This method correctly specifies 64.69% of firms[83]. The significance of this is that it might be useful to focus on the order of the questions when analysing responses. If later questions are based on earlier ones then adding more weight to the first set of questions might be more useful in finding out the level of malfeasance.

One outcome of this analysis is that the responses of firms who have said "yes" more often than not are treated with more confidence than firms who say "no" more often. Therefore, despite the possibility that by saying yes most of the time a firm is potentially admitting to graft and malpractice, they are treated as "honest" in the framework of the RR questions. There is a question of whether or not these responses can be trusted. By answering a higher amount of questions with a yes these firms are effectively displaying their honesty and their corrupt nature at the same time. Previous literature has questioned whether reticent respondents can bias firm-level surveys; another possible question is whether truth-telling corrupt firms might bias these surveys as well. If reticent managers (RR liars) might bias sensitive surveys then non-reticent ("honest") but corrupt firms might seek to hide the truth when the questions do not have the protection of the RR design.

4.5.5 The Order Of Responses For The Randomised Response Questions

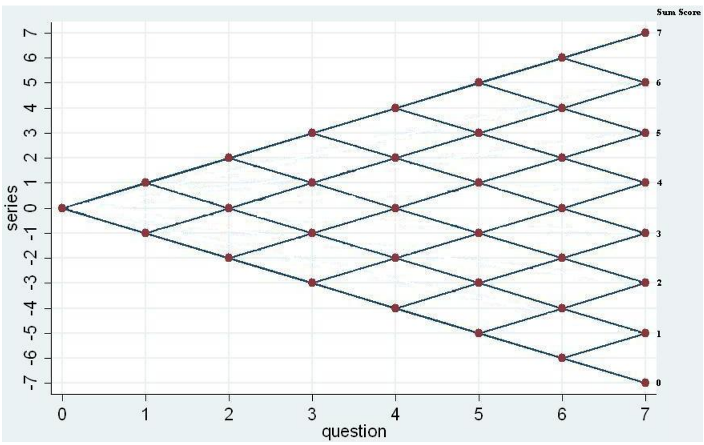
Another method to detect strategic responses with the RR questions is to look at the order in which the responses were given. If a set of seemingly reticent responses were quite common this might give evidence for systematic underreporting by the respondents. Looking at the order of responses potentially adds another dimension

[83]The corresponding values for the logit estimations for the respective models are: 64.16%; 64.74%; and 64.58%

to the data. For example, interviewees might have been honest for the first set of questions, but, as the questions continued, might have wanted to avoid seeming corrupt or looking bad and so reported “no” for the rest of the questions. Such an interviewees set of responses might look something like: {1,1,1,1,0,0,0}. Looking at the order of responses allows one to potentially distinguish such an interviewee from somebody who reported the same number of yeses but was being honest in their responses. Somebody who was being honest, yet responded “yes” to the same number of questions might have had a set of responses similar to: {1,0,1,0,1,0,1}.

Given 7 questions and 2 possible answers to each question (yes or no) there are $2^7 = 128$ possible types of responses. One can illustrate the sets of all possible yes/no combinations that give rise to a given sum score using a diagram that tracks each firms set of responses, giving a sum of “+1” for every yes and “-1” for every no. This is illustrated in Figure 10. In this figure, the connecting points represent the questions, the paths heading upwards represents questions answered with a “yes” and the paths heading downwards represent the questions that were answered with a “no”. The numbers on the far right hand side of the figure represent the sum scores. As previously mentioned, two firms can have a different set of responses yet still have the same sum score, and therefore arrive at the same final point.

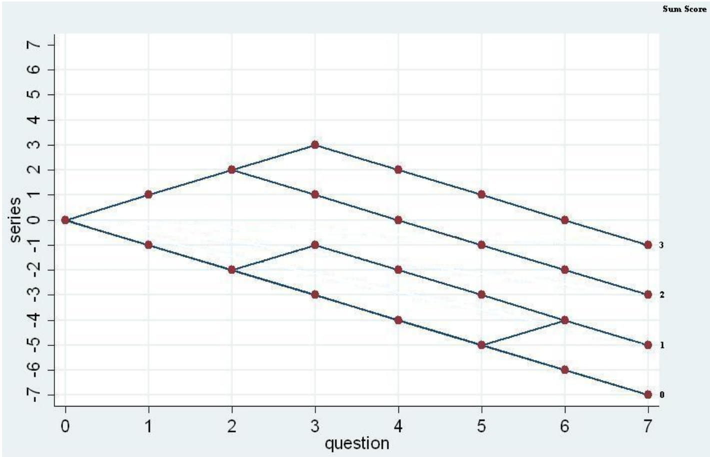
Figure 10: Potential Response “Paths” To 7 Randomised Response Questions



In the data there are five set of response combinations (paths) that occur on more that 50 occasions. These are: {0,0,0,0,0,0,0}[84]; {0,0,0,0,0,1,0}[85]; {1,1,0,0,0,0,0}[86]; {1,1,1,0,0,0,0}[87]; and {0,0,1,0,0,0,0}[88]. These patterns are illustrated in Figure 11.

[84] 414 Firms (13% of the sample) have this response pattern.
[85] 88 firms (3%) have this response pattern.
[86] 83 firms (3%) have this response pattern.
[87] 57 firms (2%) have this response pattern.
[88] 55 firms (2%) have this response pattern.

Figure 11: Most Common Response Patterns To The 7 Randomised Response Questions



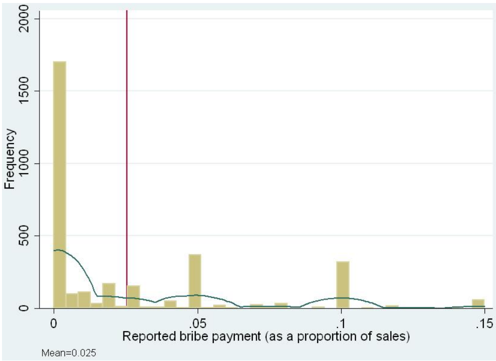
The most common patterns displayed in this figure show that amongst these firms "no" responses are more common than "yes" responses. Hence the downward trend in the graph and the sum scores being equal to 3 or less. This is significant because it can aid future research in RR questioning to test if the order of questions is relevant for the trend of yeses and nos.

4.5.6 Analysis Of Peer Group Reporting

The nature of the indirect questions might raise the issue of who the firms were referring to when they gave their responses: themselves or their peers. If the firm was referring to itself, then the answer can be interpreted in relation to that firm, however, if the firm was referring to other establishments, then the answer might be interpreted as an average figure for a group of similar firms. Figures 12 & 13 show that the responses from the two indirect questions that ask for bribe figures are clustered around 0%, 5%, 10% and 15%. This relationship remains when the sample is disaggregated by size-sector-location cells.

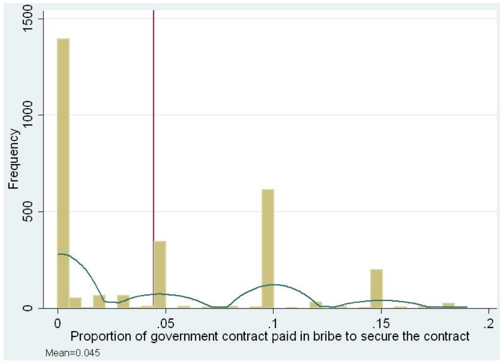
Table 4.14 presents an analysis of variance for the proportion of total sales paid in informal gifts/payments. The overall test of firm size is not significant at the 10% level (F=1.93, p=0.146). Therefore, after adjusting for industry and location, we cannot reject the null hypothesis that the average bribe (as a proportion of sales) is equal amongst the different firm sizes. The average bribe seems to vary between sectors and regions. Table 4.15 shows that the adjusted mean bribe value for small firms is 2.5% of total annual sales. Figure 14 shows these adjusted means with their repective 95% confidence intervals. Much of the variation in bribe

Figure 12: Kernel Density Estimate Of The Proportion Of Sales Paid In As Informal Payment/Gifts To Speed Regulatory Processes



N=3,200

Figure 13: Kernel Density Estimate Of The Proportion Of Government Contract Paid In Informal Payment/Gifts To Secure Contract



N=2,865

payments amongst firms of different sizes comes from the difference between small firms and medium sized firms, this is shown in Table 4.16 which shows the results from pairwise and joint chi-squared tests of equality for each respective group, controlling for location and industry. The figures in this table are derived from pairwise comparisons of the average bribe (as a proportion of sales) conditional on being in each size category: small; medium; and large. Results from this table indicate that at the 5% level the average predicted bribe (proportion) between small firms and medium sized firms; small and large; and medium and large firms, are not significantly different from one another. This suggests that bribe amounts (in Naira) increase with firm size and that this variation supports the argument that the firms are reporting about themselves.

Table 4.14: ANOVA For Informal Gifts, Controlling for Sector, Size, and Location

	Number of obs = 3200			R-Squared=0.038	
	Root MSE=0.037			Adjusted R-Squared=0.037	
Source	Partial SS	d.f	MS	F	Prob>F
Model	0.171	5	0.034	25.41	0.000
Size	0.005	2	0.003	1.93	0.146
Sector	0.008	2	0.004	3.11	0.045
Region	0.161	1	0.162	119.86	0.000
Residual	4.309	3194	0.001		
Total	4.481	3199	0.001		

Table 4.15: Adjusted Means Of Bribery, By Firm Size, Controlling for Sector And Region

	Margin	Delta-method Std. Err.	z	P>|z|	95% Confidence Interval	
Firm size						
Small	0.025	0.001	33.98	0.000	0.023	0.026
Medium	0.028	0.001	19.30	0.000	0.025	0.031
Large	0.021	0.011	1.90	0.057	-0.001	0.043

Table 4.16: Contrasts Of Predictive Margins

	Contrast	Delta-method Std. Err.	95% Confidence Interval		d.f	Chi2	P>chi2
Firm size							
Small vs. Medium	-0.003	0.002	-0.006	0.000	1	3.69	0.0547
Small vs. Large	0.004	0.011	-0.018	0.026	1	0.11	0.736
Medium vs. Large	0.007	0.111	-0.015	0.029	1	0.38	0.537
Joint					2	3.85	0.146

4.6 Conclusion

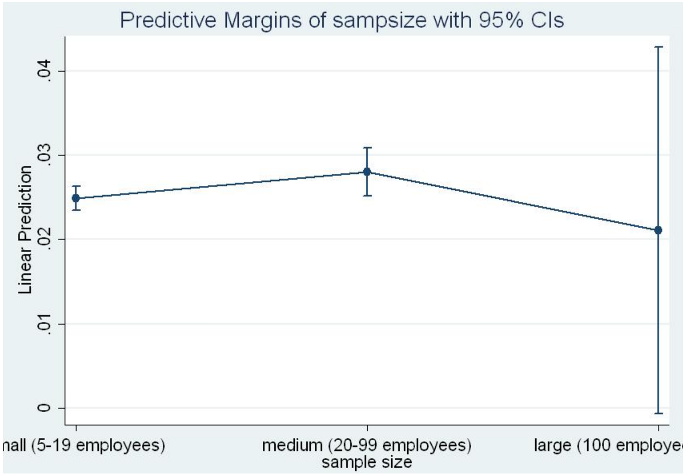
Despite the significant influence that corruption is believed to have in much of the developing world, the difficulty in directly observing corrupt acts has led to an absence of evidence, and consequently consensus, on the most effective ways to reduce its potential adverse impact on an economy. The accuracy of corruption measures is important because of their use by academics and institutions to measure corrupt behaviour around the world [Olken , 2009]. Corruption measures such as Transparency International's Corruption Perceptions Index (CPI) [Lambsdorff , 2003] and the World Bank's Governance Indicators [Kaufmann, Kraay & Mastruzzi , 2005] are formed on the basis of corruption perceptions data and are used to asses governance at the national and sub-national levels. Such data has also been used to discuss the determinants of corruption[89].

The current study set out to discuss some of the methods that have been used in the economics literature to get honest answers from sensitive questions, including questions about personal experiences with crime and corruption. Some of the literature concerning such issues use direct methods of asking businesses and individuals about their behaviour with regard to the tabooed acts. However, it has been recognised that such methods of data collection are subject to misreporting bias. Other techniques have been developed in order to improve the data on sensitive topics by reducing the guilt or taboo associated with such acts. A number of survey methods have been discussed, including different specific question strategies used to get more honest responses. The wording of questions was also raised as a potential issue of importance.

This study identified two main strategies that have been applied to the study

[89] A survey of this literature can be found in Rose-Ackerman [2004].

Figure 14: Adjusted Mean Of Informal Payment, By Firm Size



of corruption and illegal activities. These are: indirect questioning (IQ); and randomised response questioning (RRQ). Both methods aim to get more positive answers by reducing the guilt associated with saying 'Yes" in response to a positive question. The indirect methods that were discussed do this by asking a question about somebody who is similar to the interviewee, rather than the interviewee themselves. This is meant to allow the respondent to admit to performing an illegal act whilst possibly maintaining that they were referring to somebody else. The taboo of affirming a trait that is socially looked down upon is meant to be reduced because only the interviewee knows who he/she is referring to. The RR method seeks to reduce the taboo associated with similar questions by allowing the interviewees answer to depend on a random event, the result of which is only known to the interviewee, Therefore, similar to IQ, RRQ allows the interviewee to say "Yes" without necessarily admitting to an undesirable trait.

This chapter put both methods to the test by analysing the results of a firm-level survey that used both methods to ask similar sensitive questions to managers in Nigeria. Results from section 4.5 were in favour of indirect methods as a means of getting honest responses to questions about bribery and tax avoidance. While the highest RR estimate for the prevalence of a sensitive act came to 3.6% of the sample, the lowerst indirect estimate for any sensitive trait was 53%. Possible reasons for the relative failure of the RR technique might be the design of the procedure. Another set of questions using a randomising device with a lower probability of selecting the sensitve question might possibly yield a higher rate of admission. However, this would also lead to a decreased efficiency in any estimate of transgression.

The distribution of responses showed a positively skewed graph that tailed out to the right. The presence of reticence (the unwillingness of some interviewees to respond positively to a sensitive question even when required to do so by the RR technique) was given as a reason for the skewed distribution and for the much higher number of firms with a sum score of zero. This situation suggests two separate distributions, one for those who follow the rules and one for those who do not (Van Den Hout, 2012)[90].

This study added to the literature on the RR by developing a model of reticence and the response to RR questions. This model allowed for the calculation of a lower bound estimate of reticence to be made. This lower bound was based on the assumption that nobody committed any of the acts in question and came to a value of 16.9%. This study was also able to generate a prediction of the response to the final question of a series of RR questions based on the responses to the previous questions and a set of explanatory variables. This model was able to correctly predict the final response for 64.16% cases.

Finally, an analysis of peer group behaviour was conducted. The reason behind this was to dig deeper into who the firms were referring to when answering the indirect questions. Size, sector, and regional comparisons were made with the result that after controlling for sector and region, relatively little variation in bribe payments was due to firm size. The variation that did occur between firms of different sizes were mostly due to the differences between small and medium sized firms.

This work has provided evidence for the superiority of indirect questioning over randomised response questioning in getting honest responses from sensitive survey questions. Previous work has studied the effectiveness of randomised response questioning over direct questioning. Further work can seek to simultaneously study the point estimates and efficiency of the direct questioning, indirect questioning and randomised response questioning methods.

Despite these conclusions, this investigation has raised the question of why firm managers continue to be reticent, when asked RR questions, despite being given anonymity. The next chapter investigates possible reasons for this phenomenon.

[90] Correspondence with author.

5 What Drives Reticence? Reporting Bias From Monopolies And Distrustful Firm Managers

Abstract

This chapter examines the determinants of underreporting (reticence) on randomised response questions. By looking at the relationship between firm-specific reticence and other firm-specific and industry-location specific variables it is found that (mis)trust, indicated by the proportion of contracts arranged before delivery, is a significant predictor of reticence. This seems to suggest that firm managers who are more cautious in their business dealings are also more cautious with the randomised response technique. In such cases, weighted estimates of the prevalence of sensitive traits might be derived without the use of the RR technique but through the use of variables relating to the nature of firm-level contracts.

5.1 Introduction

The previous chapter compared the indirect questioning and randomised response questioning methods against each other to see which one is preferred when asking sensitive questions. In doing so, another problem was raised: why do firm managers exhibit reticence when asked sensitive questions via randomised response despite being given assured anonymity? The current chapter seeks to answer this question.

This study investigates the reasons for underreporting of sensitive acts by firms, even when those same firms have protection against being found guilty of those acts. The role of trust and the perceived risk of detection is examined and found to have significant effects on the propensity to report truthfully. The results suggest that some of the factors which affect the nature in which companies engage in economic activity might also affect the exchange of sensitive information.

A number of reasons have been provided for the lack of reliability of survey data. These include: social desirability bias; cognitive problems; and the effect of attitudes on responses. In the previous chapter, social desirability bias was provided as one of the main reasons for the underreporting of questions about corruption. Interviewees generally want to avoid looking bad in front of the interviewer[91]. Such bias is of special importance when the interviewer; research organisation; or framework[92] can be associated with the questions in the survey. For example, reported prejudice can decrease when a survey is administered by an ethnic minority; and might increase if interviewees believe that they are being psychologically monitored for lie-detection [Bertrand & Mullainathan , 2001]. Also, bribery is likely to be underreported if the survey is conducted by an anti-corruption agency and if the interviewees are aware of this. The current study seeks to overcome this potential problem by using data that was not acquired by an anti-corruption agency. Also, the questions concerning corruption are towards the end of the survey when the interviewer would have gained some rapport with the interviewee [Svensson , 2003]. Furthermore, a few questions concerning crime and corruption are scattered throughout the survey, therefore, the topic is unlikely to come as a shock to the interviewee.

On the other hand, some research has uncovered the effect of cognitive processes on the way that interviewees respond to survey questions[93]. Findings from this literature suggest that simple changes can significantly alter the processing and manipulation of survey questions. One such change is the altering of the order

[91] Some studies have tried to reduce this effect by increasing the physical and social distance between the interviewer and the interviewee [Fischbacher & Heusi , 2008, Abeler, Becker & Falk , 2012].

[92] For example, the purpose of the study.

[93] Experimental evidence on the effect of cognitive processes on survey responses can be found in Tanue [1992] and Sudman, Bradburn & Schwarz [1996].

of questions. Changing the order might affect the responses to questions because people try to provide answers that are consistent with their previous answers in the survey. Another potential issue is that previous questions might stir up attitudes or memories which can affect the responses to future questions. A further channel through which cognitive processes can affect responses to questions is through question wording. One example [Rugg , 1941] relates to the respective responses of comparable cross-sections of the public to the two following questions:

- Do you think that the United States Of America should allow public speeches against democracy?
- Do you think that the United States Of America should forbid public speeches against democracy?

Whilst three-quarters of the first sample of respondents reported "No" to the first question; just over half of the respondents in the second sample reported "Yes" to the second question. In this case it seems to be a framaing effect that causes the difference in responses to the two respective questions rather than an ordering effect that might occur if both questions, in different orders, were asked to both samples. The questions asked to the firm managers in this chapter were asked in the same order and had the same wording for all interviewees, suggesting that responses can be comparable across managers. Problems can also occur with different scales, for example, when given the choice between reporting bribes as a percentage of sales and in local currency value, reporters in local currency tend to report lower amounts, when converted, than those who report as a percentage of sales [Clarke , 2011]. Also, when asked how much television they watch per day, interviewees respond differently depending on the size of the first category and the size of the subsequent increments for each category. The RR questions in this study do not have this problem since they require a dichotomous (yes/no) response. Finally, there might be a problem if interviewees do not use much mental exertion in answering the questions. When given a list of responses interviewees often pick the first and/or the last option in the list. The questions in this study do not require a list so this problem should not occur here. The absence of attitudes, sometimes respondents think that they should report having an opinion on something when questioned about it; the confusion over the specific attitude that one holds and/or why one holds that attitude; and the presence of conflicting attitudes can also affect the way that people respond to survey questions. These should not be a problem in this study since the questions used concern actions and not attitudes.

5.2 Framework

This section outlines the framework used to study the determinants of reticence. The following subsections correspond with the analysis conducted by Sah [2007] by treating a firm's behaviour as being determined by previous experiences; firms use the state of affairs in the previous period to inform their decisions in the present period.

5.2.1 Managers (Agents)

Consider an economy consisting of firm managers who interact with each other in order to purchase and sell goods and/or services. Each manager has a level of (dis)trust, r, for every other manager in the economy. Trust is based upon the managers' previous dealings in the economy with other managers. Trust also affects how the manager will engage in business in the future. Managers are randomly allocated to each other in each period, no manager knows who they will meet in the next period. The choice of how much to trust the next manager is made by each manager in each period and depends on his/her estimate of the mean proportion of trustworthy managers in the economy.

5.2.2 Interviewers (Principals)

In this economy, there exists a predetermined number of interviewers. Interviewers are employed by a research institute for the purpose of acquiring information from the managers. All interviewers are employed by the same research institute. Researchers employ the interviewers with the purpose of acquiring information from the managers. Interviewers would, therefore, like the managers to reveal some potentially sensitive information to the interviewer.

5.2.3 Two Scenarios

This study considers two potential scenarios:

1. Firm managers see the questions up front and know how many questions are going to be asked. In this case they choose s (the number of yeses) to maximise their utility from remaining in business and continuing to receive the stipend provided by future interviews. This will be dependent upon re-interview and whether or not the information is leaked.

2. Managers do not know how many questions they will be asked. In this case they play the game (tell truth/lie) after being asked each question, and the equilibrium evolves: either the optimal strategy is to tell truth or lie, in which case managers display honest and reticent behaviour; or to play a mixed strategy, in which case managers display behaviour identified as possibly candid/possibly reticent.

5.2.4 Meetings Between Interviewers And Managers

At time $t_{int} > 0$, managers are randomly matched with 1 interviewer in a (potentially) one-shot game. t_{int} can vary for different managers. At time $t_{int} > 0$ each interviewer is matched with 1 manager, however, interviewers are matched with more than one manager over the course of the total time frame, $t_1 ... t_T$, where T is the final time period. This model allows for the game to go on forever so that $T = \infty$, it also allows for a limited time frame, $T = c < \infty$.

Therefore, interviewers are matched with $m > 1$ managers. Interviewers want the managers to reveal some potentially sensitive information to the interviewer. When matched to the interviewer, each manager has a choice of one of two options: reveal the truth, or lie. Revealing the truth could lead to, with a very small probability, G, the information being leaked and used against the manager. However, this probability is not known to the managers, although managers have an idea about the value of this probability, $g(r)$. Managers are not certain about their evaluation of the probability of leakage. The managers' idea about the probability of leakage, $g(r)$, is a range from a minimum probability, $\underline{g(r)}$, to a maximum probability, $\overline{g(r)}$. $\underline{g(r)}$ and $\overline{g(r)}$ are known to the managers. The size of the range, $g^* \equiv \overline{g(r)} - \underline{g(r)}$ remains the same for all managers. However, the position of $g(r)$ on the $[0, 1]$ interval is allowed to vary amongst managers. The lower bound of $\underline{g(r)}$ is 0 and the upper bound of $\overline{g(r)}$ is 1. Therefore, $\underline{g(r)}$ can take a value within the interval $[0, \overline{g(r)} - g^*]$ and $\overline{g(r)}$ can take a value within the interval $[\underline{g(r)} + g^*, 1]$.

Therefore:

$$g_i^* = g^* \quad for \quad all \quad i$$

$$\underline{g(r)} \in [0, \overline{g(r)} - g^*]$$

$$\overline{g(r)} \in [\underline{g(r)} + g^*, 1]$$

The managers' actions in this game depends on their level of trust, r, at time t_{int}.

If the true probability of leakage, G, is on or above a threshold, then the information will be leaked to the authorities. If it is below this threshold, then the information will not be leaked. For simplicity, let this threshold be 0.5, and let this information be known to all managers. Doing this makes the problem one of discerning whether or not the probability of detection, G, is greater than or less than a half.

Since the managers do not know the probability of detection, but only have an estimate about the range that it can take, $g(r) = g^* \equiv \overline{g(r)} - \underline{g(r)}$, then the managers' problem is to decide whether the range lies above the threshold,

$\underline{g(r)} > 0.5$, or below the threshold, $\overline{g(r)} < 0.5$. In the first case, if $\underline{g(r)} > 0.5$, the manager believes that the probability of leakage, G, is greater than a half and the information will be leaked. In the second case where $\overline{g(r)} < 0.5$, the manager believes that the probability of leakage is less than a half and that the information will be kept secret.

In the case where part of the range lies above the threshold and part of it lies below the threshold, the larger portion of the range will carry the most weight for the manager's decision. So if $\overline{g(r)} - 0.5 > 0.5 - \underline{g(r)}$ the manager believes that the probability of leakage, G, is greater than a half and that the information will be leaked. Alternatively, if $\overline{g(r)} - 0.5 < 0.5 - \underline{g(r)}$, the manager believes that the probability of detection is less than 0.5 and the information will not be leaked.

In the limiting case where $\overline{g(r)} - 0.5 = 0.5 - \underline{g(r)}$, the manager believes that the probability of detection is equal to a half and that the information will be leaked. This setup is equivalent to each individual manager taking the expected value of $g(r)$ to be $\frac{\overline{g(r)}+\underline{g(r)}}{2}$ and using this to make a judgement about the value of G.

If the information is leaked to the authorities. Managers face a punishment cost, $c(s)$, where s represents the number of yes responses that the manager gives. This study posits that $\frac{\delta c(s)}{s} > 0$, so that a higher admission to corruption will earn a more severe punishment from the authorities.

The firm managers are given a payment, p_t, for participating in this game, regardless of their actions. At the end of the game, the interviewers judge how truthful they believed the managers' responses were. If the responses are judged to be sufficiently truthful, then the interviewer will allow the manager to be interviewed again at some time in the future, in which case the manager would receive payment, p_{t+1}. If the interviewer believes that the manager gave false answers in the game, i.e. that the manager did not give sufficiently truthful responses, then the interviewer will report this to the research institute, who will not allow the firm to be interviewed again, in which case their future payoff from the interview process is zero, $\sum_{t=int+1}^{\infty} p_t = 0$.

The firm managers have to maximise the utility of revealing this sensitive information subject to the constraint that: 1) The information might be leaked and their (potential) guilt might be revealed. If this information is revealed and the firm is guilty of corruption, then the manager faces a cost of punishment, $c(s)$. 2) Perceived lying would result in the dissaproval of the interviewer and would result in a reduced probability of the manager being approached for an interview again[94]. If this is the case, the firm would lose out on a payment for participation, $p_t(s(g(r)))$, where p is the present discounted value of the future payments for future interviews.

[94] Alternatively, the probability of a future interview could be reduced to 0 by placing the manager on a "blocked list".

When each individual manager maximises his/her utility subject to these constraints, three types of individuals/actions follow:

1. Individuals who are driven by the probability that their actions will be revealed. These choose to lie despite the potential lost future income. For this to occur(c(s)) must be greater than p(s(g(r))).

2. Individuals who are driven by the potential future income that they will receive from telling the truth. These will choose to tell the truth despite the probability of leakage. For this to occur(c(s)) must be smaller than p(s(g(r))).

3. Individuals who face a flat region in their mazimisation process. So that neither constraint is binding. For this to occur(c(s)) must be equal to p(s(g(r))).

5.2.5 Possibility Of Re-Interview

The possibility of another interview, and by implication, another chance to earn the stipend for taking part in the interview process, is related to the (dis)approval of the interviewer. This disapproval is a function of the binomial distribution, which the responses are related to. Since the expected number of yeses under the angels assumptions is 3.5, then any number below this value suggests lying behavior and will attract disapproval from the interviewer, decreasing the chance of a future interview.

In order to allow for more flexibility, this study takes a continuous approach to the decision for a re-interview. The expected future payoff depends on the probability of re-interview, which is a function of the observed number of yeses and the expected number of yeses: $p_{t+1}(s, E(s)) = f(Pr(Re - Interview)) = f(M, r, \theta, b)$. Where M is the number of RR questions, $1 - \theta$ is the probability of being required to answer the RR question truthfully, b is the nature of the manager's level of guilt. In the current setup, this simplifies to: $p_{t+1}(s, E(s)) = f(Pr(Re - Interview)) = f(7, r, 0.5, b)$.

The probabilitiy of being re-interviewed increases with the level of guilt and decreases with the level of reticence. More formally: $p_r < 0$; $p_b > 0$; and $p_{br} = p_{rb} < 0$. This means that reticence has a greater impact on future payoff than guilt[95].

Since $E(s) = 3.5$ is the benchmark by which a dishonest response is identified, this is incorporated into the re-interview function to give:

$$Re - Interview = \begin{cases} 0 < Pr(Re - Interview) \leq 1 & s \geq 3.5; \\ 0 & s < 3.5 \end{cases}$$

[95]This result was shown in the previous chapter.

5.2.6 Leakage Of Information

The probability of leakage of information is calculated by the manager. This probability is positively related to the number of yeses. In other words, the manager believes that if they respond yes to a relatively large number of questions, then the research institute will reveal it to the authorities: $Pr(Leak) = f(\gamma_0, \gamma_1 \cdot s)$. Where γ_0 represents the belief, by the managers, that information might be leaked regardless of the answers provided; and γ_1 represents the part of the decision to leak that is based on the answers to the RR questions. From the previous discussion, it is believed that $\gamma_0 \geq 0$ and $\gamma_1 > 0$. The manager chooses the number of positive responses, s, in order to minimise the probability of leakage, $Pr(leak)$. This analysis informs the empirical section to come by creating expressions which can be used to represent the latent trust in the interviewer and the desire not to be seen as socially bad. The term γ_0 can be interpreted as the latent trust in the interviewer not to leak the answers independent of the responses given. On the other hand, the term γ_1 is more closely linked to the desire not to appear bad to the interviewer, and the contribution of this term to the probability of leakage depends on the responses of the manager. Whereas Sah [2007] modelled the firm's business deals being determined by previous deals with other firms; the current study models meetings between the manager and interviewer being determined by a firm's previous business deals and its experience of the business environment.

5.3 Data & Construction Of Variables

This subsection describes the data used in this study. Descriptions of the variables used in the analysis are provided in subsection A.3. Summary statistics for the data are provided in subsection A.3.1. The data used in this chapter comes from the Enterprise survey which was described in Section 2.4. The firm-specific variables come from the questionnaire that was used in the survey; the information concerning industry-location population is taken from the sample survey methodology [World Bank , 2007, Iarossi, Mousley & Radwan , 2009].

5.3.1 Trust And Experience In The Business Environment

This study uses three variables as the main indicators of trust. These are: the percentage of total orders that are written, as opposed to oral with a witness and oral without a witness; the percentage of material orders that are paid for after delivery, as opposed to on delivery or before delivery; and the percentage of sales orders that are paid for before delivery, as opposed to on delivery or after delivery. These variables are chosen because they seem to represent things that a non-trusting company should care about.

This study also uses a dummy variable for whether or not a company subcontracted any part of its production process to another company. Subcontracting work to another company arguably requires a level of trust that the company will perform the work and will do it to a certain standard. Also, the number of years that the company has known the primary supplier for its main input is used as an explanatory variable. This is meant to denote the level of mistrust. The years of experience of the senior manager and its squared term are also used to see the effect of business experience on reticence.

5.3.2 Level of Anonymity And The Fear Of Detection

One reason for reticence might be the fear of detection. Reticent companies might, arguably irrationally, reason that the answers from the RR procedure might be used against them somehow. If so, since the surveys were carried out anonymously, the possibility of detection would be related to the possibility that the firm could be identified by observables. One of these is the number (population) of other companies in the industry-location cell. If a company is the only chemical manufacturing business in Bauchi state, then it might fear that this fact might make it identifiable. On the other hand, if a company is one out of a hundred wood manufacturing businesses in Lagos, then it might be more at ease with answering sensitive questions, since the possibility of detection, as described above, would be reduced.

The preceding analysis suggests the use of the number (population) of companies in the industry-location cell as a variable to measure the level of identifiability of the company. The preceding analysis suggests that the larger the population of companies, the lower the propensity to be reticent. This study also includes the number of close competitors, as reported by the company itself, as an explanatory variable. Due to the fact that this variable is self-reported by the company, it might be more closely related to reticence than the total population of firms. The total population of firms is, arguably, less known to the individual company than the number of close competitors. Also, the total population in the industry location cell might include somewhat different companies in terms of both product and location. Another argument in favour of using subjective reports of the number of close competitors is that this question was asked before the randomised response questions and also before the questions about corruption; therefore there is no danger of the managers using their perception to justify their behaviour.

5.4 Empirical Methodology

5.4.1 Misclassification Of Reticence

The definition of reticence is a manager that answers “No” with a positive probability when they are supposed to answer “Yes”, when doing so might be interpreted as them having committed a socially undesirable act [Azfar & Murrell , 2009, Clausen, Kraay & Murrell , 2010]. This definition does not require that reticent managers always give untruthful response, but that they do so on at least some occasions. In practice, the measure of reticence that is used is the number of “yeses” on a series of randomised response questions. A manager with zero yeses is classified as reticent, a manager with at least one yes is classified as possibly candid.

One potential problem with trying to find the causes of reticence is that the traditional measure of reticence might be misclassified because of misreporting errors. Labelling all firms who answer “No” all the time as reticent and the rest of the sample as possibly candid potentially misclassifies some of the reticent as possibly candid. This can be described using the latent variable specification of a binary outcome model [McFadden , 1984, Greene , 2003] and the notation of Hausman, Abrevaya, Scott-Morton [1998], where y_i^{true} is a latent variable, and i ranges from 1 to the sample size, N. The latent variable can be described by:

$$y_i^{true} = x_i'\beta + \epsilon_i$$

where ϵ_i is an independent and identically distributed error term. The observed response can be represented as:

$$y_i^{observed} = 1(y_i^{true} \geq 0)$$

where $y_i^{observed}$ is the reported answer; and 1(E) represents the indicator function that is equal to 1 if E is true and zero if E is false. In the absence of misclassification of the binary variable the observed response is also the true response.

This study focuses on the type of misclassification in which the misclassification error depends on the true response, y_i^{true}, but is independent of the explanatory variables, x_i.

The definition of reticence used previously can be represented as:

$$r = \begin{cases} 1 & \text{if } s = 0; \\ 0 & \text{if } s > 0. \end{cases}$$

where r represents reticence and s is the number of yeses to a series of randomised response questions. This criterion has the possibility that some firms are misclassified as 0 (possibly candid) when in fact they should be 1 (reticent). In such a case where some 0s should be 1s: $\alpha_1 = Pr(y_i^{observed} = 0|y_i^{true} = 1) \neq 0$, where α_1 is the false-negative misclassification error. Due to the relatively sensitive nature of the questions and the relatively low probability of getting 7 tails from 7 coin flips and being innocent of all acts, this study argues that the alternative misclassification error is not significantly different from zero: $\alpha_0 = Pr(y_i^{observed} = 1|y_i^{true} = 0)$. Despite this, the false-positive misclassification error, α_0, will be incorporated into the analysis and tested for significance.

In the standard case of miclassification of a binary dependent variable, the expected value of the observed dependent variable is:

$$E(y_i^{observed} = 1|x_i) = Pr(y_i^{observed} = 1|x_i) = \alpha_0 + (1 - \alpha_0 - \alpha_1)F(x_i^{'}\beta) \quad (5.4.1.1)$$

Where F is the cumulative distribution function of ϵ_i. The probability of observing a zero is given by: $Pr(y_i^{observed} = 0|x_i) = (1-\alpha_0)+(\alpha_0+\alpha_1-1)F(x_i^{'}\beta)$. This collapses to the usual respective expressions: $F(x_i^{'}\beta)$ and $1 - F(x_i^{'}\beta)$, when there is no misclassification error in the binary dependent variable.

In the present setup, the measurement error is negatively correlated with the accurately measured variable, therefore, the classification error will lead to a downward bias in the estimates of the effect of x on y^{true} [Bound, Brown & Mathiowetz , 2001]. To see this, note that the marginal effect of an explanatory variable on the observed response is:

$$\frac{\delta Pr(y_i^{observed} = 1|x_i)}{\delta x} = (1 - \alpha_0 - \alpha_1)f(x^{'}\beta)\beta \quad (5.4.1.2)$$

which is always less than the marginal effect on the true response: $f(x^{'}\beta)\beta$. Moreover, the marginal effect on the observed response always differs from the marginal effect on the true response by a factor of $(1 - \alpha_0 - \alpha_1)$, no matter at what value of x the marginal effect is evaluated.

The parameters α_0; α_1; and β can be estimated via the maximum likelihood method by maximising:

$$\begin{aligned} lnL =& \frac{1}{n}\sum_{i=1}^{N} y_i^{observed} ln(\alpha_0 + (1 - \alpha_0 - \alpha_1)F(x_i^{'}\beta)) \\ &+ (1 - y_i^{observed})ln((1 - \alpha_0) + (\alpha_0 + \alpha_1 - 1)F(x_i^{'}\beta)) \end{aligned} \quad (5.4.1.3)$$

with respect to α_0; α_1; and β.

An additional condition is required for identification of the parameters in the model. This condition, the monotonicity condition, states that the sum of the misclassification errors must be less than 1.
Monotonicity Condition (MC1): $\alpha_0 + \alpha_1 < 1$.

This can be seen when considering a symmetric function, F, where $F(d) = 1 - F(-d))$. Defining: $\tilde{\alpha_0} = 1 - \alpha_1$, $\tilde{\alpha_1} = 1 - \alpha_0$ and $\tilde{\beta} = -\beta$. Then:

$$\begin{aligned}
\tilde{\alpha_0} + (1 - \tilde{\alpha_0} - \tilde{\alpha_1})F(x_i^{'} - \tilde{\beta}) =& 1 - \alpha_1 + (1 - (1 - \alpha_1) - (1 - \alpha_0))(1 - F(x_i^{'}\beta)) \\
=& 1 - \alpha_1 + (1 - 1 + \alpha_1 - 1 + \alpha_0)(1 - F(x_i^{'}\beta)) \\
=& 1 - \alpha_1 + (\alpha_1 + \alpha_0 - 1)(1 - F(x_i^{'}\beta)) \\
=& 1 - \alpha_1 + \alpha_1 + \alpha_0 - 1 - (\alpha_1 + \alpha_0 - 1)F(x_i^{'}\beta) \\
=& \alpha_0 + (1 - \alpha_0 - \alpha_1)F(x_i^{'}\beta)
\end{aligned} \tag{5.4.1.4}$$

Therefore, any estimators based on Equation 5.4.1.1, such as maximum likelihood and non-linear least squares, cannot distinguish between $(\alpha_0,\ \alpha_1,\ \beta)$ and $(1 - \alpha_1,\ 1 - \alpha_0,\ -\beta)$. MC1 rules out this possibility since if $\alpha_0 + \alpha_1 < 1$, then $(1 - \alpha_1) + (1 - \alpha_0) > 1$.

Since this study argues that α_0 is not significantly different from zero, then the MC1 condition implies that $\alpha_1 < 1$, which appears to be reasonable. The only way in which this could be violated is if $\alpha_1 = 1$ which does not agree with the data since the actual distribution of "yeses" shows that more yeses are observed than expected under the angels assumption, therefore, at least some respondents appear to be answering truthfully. Using this result, it appears that the mononicity condition is satisfied in this study and the parameters $(\alpha_0, \alpha_1, \beta)$ can be identified using maximum likelihood estimation.

5.4.2 Possible Extensions

The previous subsection argued that since the 0/1 measure of reticence might misclassify some firms as possibly candid when they were indeed reticent, this might cause a source of bias. One way to control for this bias is to estimate the misclassification errors, α_0 and α_1, along with the coefficient of the model, β. Another potential solution is to use the number of yeses, 0 to 7, as a dependent variable instead of the binary measure of reticence. This would transform the model from a binary outcome model to a multiple outcome model. Doing this would potentially increase the amount of information used in the estimation and would bypass the problem of misclassifying some reticent managers as possibly candid. To be sure, there still exists the possibility that some of the firms were lying, so somebody who said 2 yeses should have actually stated 3 yeses. Nevertheless, to deal with the potential misclassification of the number of yeses, one can use the analysis of mismeasured discrete dependent variables used in Abrevaya & Hausman [1999], which extends the binary response framework to deal with ordered discrete choice variables with more than 2 possible values.

5.5 Empirical Results

This section looks at the empirical results of the analysis. Summary statistics from the data are discussed. Results from an econometric exercise that models the propensity to be reticent are also analysed.

5.5.1 Summary Statistics

Summary statistics for trust-related variables are shown in Table 5.1. Different sets of variables are presented within this table. The first set of variables relate to the companies' orders. Companies were asked to state the percentage of their customers' purchase orders that were: written; oral, with a witness; and oral, without any witnesses. These respective percentages add up to 100%. The second set of variables relate to the annual purchase of material inputs made by the companies. The companies were asked to state the percentage of total annual material inputs or services that were paid for: before delivery; on delivery; and after delivery. These percentages also add up to 100%. The next set of variables relate to the sales of the establishment. Companies were asked the percentage of sales that were paid for: before delivery; on delivery; and after delivery. These percentages also add up to 100%. Companies were also asked if they subcontracted any part of their production to another company. Finally, the survey asked the companies the number of years that they had known the primary supplier of the main input used in their production process.

The summary statistics in table 5.1 are presented seperately for the reticent and the possibly candid. On average, 45.3% of the purchase orders from reticent companies are written. This is in comparison to 32.1% for the possibly candid set of companies. Therefore, using this as a measure of trust, reticent firms seem to trust other companies less than possibly candid firms. The last column shows that the average ratio of written orders to oral-with-witness orders for reticent firms is 2.43. On the other hand, the same ratio for the possibly candid companies is 1.05. This tends to suggest that the reticent are more cautious in their business transactions than the possibly candid group.

Both groups receive payment for the majority of their sales either before delivery or on delivery. Only 12-13% of their payments are received after delivery. The reticent seem to know the primary supplier of their main input for a year longer than the possibly candid group.

5.5.2 Predicting Reticence

Tables 5.2 and 5.3 show probit estimations for reticence, these are the main results. The dependent variable in every model is a dummy equal to one if a company is reticent (as defined by Clausen, Kraay & Murrell [2010]) and 0 otherwise. Standard

Table 5.1: Means Of Trust-Related Variables, By Reticence

		Means		Relative Shares (Division Of Means)		Relative Shares (Division of Means, By Firm)	
Description	Variable	Possibly Candid	Reticent	Possibly Candid	Reticent	Possibly Candid	Reticent
Percentage of their customers' purchase orders that were:	written	32.1	45.3	1.10	2.53	1.05	2.43
	oral, with witness	29.3	17.9	1	1	-	-
	oral, without witness	38.6	36.8	1.32	2.06	0.59	0.87
Percentage of total annual purchases of material inputs or services that were:	paid for before delivery	28.5	38.7	0.50	0.93	0.73	1.13
	paid for on delivery	56.6	41.4	1	1	-	-
	paid for after delivery	14.9	19.9	0.26	0.48	0.50	0.55
Percentage of establishment's sales that were:	paid for before delivery	33.2	36.7	0.60	0.72	0.97	0.96
	paid for on delivery	54.9	50.7	1	1	-	-
	paid for after delivery	11.9	12.7	0.22	0.25	0.34	0.36
Percentage of establishment's total sales that came from selling:	intermediate products and services	6.9	11.1				
Did you subcontract any part of your production?		0.13	0.10				
For how many years have you known the primary supplier of the main input used?		7.07	7.70				

For each set of trust-related variables; columns 5 and 6 choose the middle category as the base category and divide the means of the other categories by the mean of the base category. The figures in columns 7 and 8 are generated by doing the same thing for each respective firm and then taking the mean of these figures. The last set of figures, in columns 7 and 8, are based on a lower sample size due to division by zero for some firms. Wilcoxon signed rank sum tests for the equality of shares (last 2 columns) reject the null hypothesis of equality for each pair of shares at the 5% level for the first pair, and at the 1% level for all other pairs.

errors are calculated using the Huber-White heteroscedasticity-consistent estimator. The first set of variables are the same as those used in Clausen, Kraay & Murrell [2010] as predictors of reticence, these are: gender of owner, age of owner, formal education of owner, industry of company, size of company, region, and wave of survey. Companies with owners who are 55 years old or above are less likely to be reticent than the base group of 46-55. Companies in manufacturing and retail are also shown to be more reticent than the base group, other, which includes: information technology, construction & transport, and hotels & restaurants. Companies located in the southern states of Nigeria are also more likely to be reticent than those in the Northern states. Also, companies that were surveyed in the second wave, showed more reticence than those who participated in the first wave. One potential reason for this is because the regions surveyed in each wave were chosen based on the state's committment to reform and the nature of its governance; therefore this result might arise because firms in states with lower perceived levels of governance have higher levels of reticence. A second potential reason is that in the second wave firm managers had been involved in an another RR exercise before the current one; this previous exercise might have generated reticence in these managers [Clausen, Kraay & Murrell , 2010].

Table 5.2: Probit Estimations For Reticence

Dependent Variable:	Dummy=1 if firm is reticent; 0 otherwise		
	6	7	8
age:≤30 (Dummy)	-0.147	-0.081	-0.116
	(0.102)	(0.164)	(0.104)
age:31-45 (Dummy)	-0.023	-0.057	-0.017
	(0.071)	(0.092)	(0.071)
age:≥55 (Dummy)	-0.198*	0.011	-0.242**
	(0.106)	(0.135)	(0.111)
manu (Dummy)	0.218***		0.176**
	(0.076)		(0.077)
retail (Dummy)	0.244**		0.211**
	(0.095)		(0.096)
south (Dummy)	0.241***	0.325***	0.242***
	(0.063)	(0.085)	(0.063)
wave2 (Dummy)	0.564***	0.894***	0.536***
	(0.068)	(0.111)	(0.068)
industry-location population	0.000	-0.001	-0.000
	(0.000)	(0.000)	(0.000)
orders_written	0.005***	0.006***	0.005***
	(0.001)	(0.001)	(0.001)
mat_paid _after_delivery	0.003***	0.009***	0.003***
	(0.001)	(0.002)	(0.001)
sales_paid _before_delivery	0	0.005***	-0.000
	(0.001)	(0.001)	(0.001)
subcontract (Dummy)		-0.118	
		(0.135)	
primary_supplier			0.014**
			(0.006)
Constant	-1.838***	-2.431***	-1.898***
	(0.148)	(0.21)	(0.153)
Other control variables	YES	YES	YES
Pseudo R-squared	0.058	0.118	0.058
Observations	3100	1699	3073
Log-Likelihood	-1132.5	-604	-1112.8
Chi-Squared	137.6	132	137.5
P Value	0	0	0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$

Dependent variable is a dummy=1 if firm is reticent, 0 otherwise. Heteroskedastic robust standard errors in parenthesis. Other control variables include: gender of owner (dummy); education level of owner (category); and size of firm (category). The coefficients on these variables are not statistically significant at the 10% level.

Table 5.3: Probit Estimations For Reticence

Dependent Variable:	Dummy=1 if firm is reticent; 0 otherwise		
	8	9	10
age:≤30 (Dummy)	0.036	0.05	0.031
	(0.17)	(0.171)	(0.170)
age:31-45 (Dummy)	-0.019	-0.009	-0.038
	(0.095)	(0.095)	(0.096)
age:≥55 (Dummy)	0.058	0.06	0.038
	(0.141)	(0.141)	(0.144)
secondary (Dummy)	0.159*	0.158*	0.162*
	(0.088)	(0.088)	(0.089)
tertiary (Dummy)	-0.157	-0.144	-0.132
	(0.277)	(0.277)	(0.276)
south (Dummy)	0.329***	0.327***	0.320***
	(0.086)	(0.086)	(0.087)
wave2 (Dummy)	0.852***	0.846***	0.823***
	(0.111)	(0.112)	(0.111)
mgr_experience	0.054***	0.054***	0.053**
	(0.02)	(0.02)	(0.020)
mgr_exp2	-0.001**	-0.001**	-0.001**
	(0.001)	(0.001)	(0.001)
competitors:1 (Dummy)	-0.05	-0.057	-0.043
	(0.319)	(0.319)	(0.319)
competitors:2-5 (Dummy)	-0.504**	-0.504**	-0.516**
	(0.203)	(0.203)	(0.204)
competitors:6+ (Dummy)	-0.875***	-0.871***	-0.887***
	(0.189)	(0.19)	(0.190)
orders_written	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)
mat_paid _after_delivery	0.007***	0.007***	0.007***
	(0.002)	(0.002)	(0.002)
sales_paid _before_delivery	0.004***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)
subcontract (Dummy)		-0.104	
		(0.137)	
primary_supplier			0.002
			(0.009)
Constant	-2.103***	-2.103***	-2.043***
	(0.309)	(0.309)	(0.312)
Other Control Variables	YES	YES	YES
Pseudo R-squared	0.149	0.149	0.146
Observations	1697	1696	1682
Log-Likelihood	-580.8	-580.5	-566.7
Chi-Squared	174.8	175.7	169.1
P Value	0	0	0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$

Dependent variable is a dummy=1 if firm is reticent, 0 otherwise. Heteroskedastic robust standard errors in parenthesis. Other control variables include: gender of owner (dummy); and size of firm (category). The coefficients on these variables are not statistically significant at the 10% level. The firms in these estimations are located in the manufacturing industry.

Turning attention to the trust-related variables, the coefficients on the three main measures of mistrust have their expected positive sign; and are statistically significant when they enter the estimations seperately in Models 3, 4 and 5. The coefficients on the percentage of written orders and the percentage of material purchases that were paid for before delivery are both significant at the 1% level. The percentage of sales that were paid for before delivery is significant at the 10% level. Furthermore, the percentage of written orders and the percentage of material inputs that were paid for before delivery both keep their sign, size and significance when all three variables enter jointly into the model (Model 6). Their size, in magnitude, is also similar to their size in the previous models where they enter seperately into the estimation. The coefficient for the subcontracting of production has its expected sign but is not statistically significant at the 10% level. The number of years that the company has known its primary supplier also has the expected sign, is positive, and is statistically significant at the 10% level. The squared term for the length of time that the primary supplier has been known did not enter significantly into the model and is omitted from this table. These results add support for the idea that companies who are less trusting in their business operations are more likely to be reticent in answering randomised response survey questions. The coefficients on all variables maintain approximately the same size and significance throughout all models; including when the variables are added sequentially.

Turning attention to the variables representing the risk of detection, the population of the company's industry-location cell had a zero effect on reticence. This effect did not change throughout the models in Table 5.2. In order to use a potentially better measure of the perceived risk of detection, the models in Table 5.3 use dummies for the number of competitors that the firm faces, as reported by the firm. These are: 1 competitor, 2-5 competitors, and more than 5 competitors. The excluded category is: no competitors. Hence, the dummy for 1 competitor is equal to 1 if the firm reported having 1 competitor, and 0 otherwise. The other dummy variables are constructed similarly.

The coefficient on the dummy for 1 competitor has its expected sign but is not significant in any of the models. Therefore, there is some evidence that, relative to the base group of no competitors, having one competitor decreases the probability of being reticent. However, this finding is not statistically significant at the 10% level. Nevertheless, the coefficients for the 2-5 and 5+ dummies both have their expected signs and are both statistically significant at the 1% level in most models. Furthermore, the coefficient on the variable for 5+ competitors is always larger, in absolute value, than the coefficient for 2-5 competitors, which, in turn, is also larger, in absolute value, than the coefficient for 1 competitor. This suggests that the more competitors in a market, the less likely a company is to be reticent.

These results provide evidence for the perceived risk of detection as a determinant of reticence amongst companies.

The three main trust variables, percentage of written orders, percentage of prepaid material orders, and percentage of postpaid sales orders, all enter sigificantly into the models in Table 5.3 and all keep their expected signs. The variables for subcontracting work and length of time that the primary supplier is known for also keep their expected signs but do not enter significantly into these models. Finally, the coefficient on the number of years of experience of the senior manager has its expected sign, is positive, and is significant in all models. The coefficient on the squared term for the years of managerial experience is negative and significant in all specifications. This suggests that an increased level of experience is initialy associated with an increased level of reticence, but that this relationship reverses after a point.

Using the information from the model with the highest Pseudo R-squared (Model 9), predicted values of the dependent variable are constructed. These values lie between 0 and 1. These values are transformed into predicted values for reticence using the following rule: if they lie below 0.5 they are converted to a 0; if they lie above or equal to 0.5 they are converted to a 1. Here, a 1 means that the firm is predicted to be reticent based on the explanatory variables, and a 0 means that the firm is predicted to be possibly candid based on the explanatory variables. Table 5.4 shows the results from this exercise. The predicted values are tabulated along with the actual values. Overall, this model is able to predict 86.4% of the actual outcomes correctly. To be sure, these results only apply to the manufacturing sector (since only firms in manufacturing were asked about their perceived number of competitors), nevertheless, this provides some evidence in favour of the models presented in this study.

Table 5.4: Comparing Actual Outcomes With Predicted Outcomes From Probit (Model 9) Estimations

Predicted	True Outcome Possibly Candid	Reticent	Total
Possibly Candid	1448	218	1666
Reticent	13	17	30
Total	1461	235	1696

5.5.3 Controlling For Interviewer And Supervisor Effects

One potential cause of observed reticence amongst companies is the interviewer. It is possible that some firms chose not to answer honestly because they did not trust the interviewer that was asking the questions. Another possibility is that the interviewer did not understand the process. This also might have affected the responses of the companies. Finally, the interviewers had supevisors to make sure that the interview was being conducted correctly. It is possible that the supervisors had an effect on the reticence of the companies. To examine whether or not the interviewers and/or their supervisors had any effect on the reticence of companies this study uses dummies for both interviewers and supervisors, respectively. Also, an interaction term is included for each interviewer-supervisor combination. These variables are only available for the second wave of the survey. Results of this analysis are shown in Table 5.5. The first column of results (Model 1) shows the coefficients from Model 9 of Table 5.3. Models 2 and 4 show some interviewer effects at the 1% and 5% levels, respectively; however, these cannot be distinguished from the wave effects mentioned earlier. When all three variables are included (Model 5), Wald Tests for the equality of the coefficients on interviewer, supervisor, and interviewer-supervisor, respectively, are conducted. These tests fail to reject the null of equality of the respective coefficients at the 10% level. This suggests that there were no statistically significant systematic effects of interviewers, supervisors, or combinations of interviewers and supervisors on the reticence of companies.

5.5.4 Controlling For Potential Political Connections

Although there seems to be no observable relationship between guilt and reticence, there might be a possible relationship between companies with political connections and reticence. One potential scenario is that a company with very strong political connections would have less reticence because they could potentially admit to engaging in any act and get away with it due to their links with powerful people. Alternatively, a company with strong ties to the government might want to hide all acts of wrongdoing so as to protect their associates in the government.

Companies in the first wave of the survey were asked to give the reason that they chose to locate in their particular state. Multiple answers to this question were allowed. Options included: it was the state of origin of the owner; the state has key natural resources that the firm uses; the state government gave concessions and benefits which made it more attractive to locate there; a location feasibility report concluded that the state was preferable to other states; and "other".

In order to test the political connections hypothesis, this study uses the response to this question as a proxy for the political connections of the company. Companies that chose to locate in a state because of concessions given by the state government are defined as being more politically connected to the state than companies who did not choose this option. The result of this exercise is shown in Model 6 of Table 5.5. There seems to be no statistically significant relationship between this measure of political connectedness and reticence.

Table 5.5: Probit Estimations For Reticence

	(1)	(2)	(3)	(4)	(5)	(6)
owner_male	0.031	0.027	0.076	0.069	-0.076	-0.135
	(0.146)	(0.199)	(0.193)	(0.203)	(0.234)	(0.245)
oage_1	0.050	0.206	0.256	0.238	0.213	0.220
	(0.171)	(0.227)	(0.222)	(0.227)	(0.267)	(0.328)
oage_2	-0.009	0.128	0.091	0.121	0.084	0.030
	(0.095)	(0.121)	(0.117)	(0.122)	(0.136)	(0.224)
oage_4	0.060	0.190	0.090	0.220	0.193	0.003
	(0.141)	(0.187)	(0.185)	(0.188)	(0.210)	(0.370)
secondary	0.158*	0.090	0.086	0.068	0.050	0.306
	(0.088)	(0.113)	(0.110)	(0.113)	(0.130)	(0.206)
tertiary	-0.144	-0.152	-0.208	-0.113	-0.265	0.653
	(0.277)	(0.365)	(0.379)	(0.352)	(0.337)	(0.459)
size_2	-0.089	0.122	0.123	0.097	-0.017	-0.240
	(0.111)	(0.144)	(0.139)	(0.148)	(0.162)	(0.245)
south	0.327***	0.056	0.025	0.068	0.125	0.711***
	(0.086)	(0.113)	(0.112)	(0.115)	(0.126)	(0.209)
wave2	0.846***					
	(0.112)					
mgr_experience	0.054***	0.059**	0.060***	0.059**	0.045	0.176**
	(0.020)	(0.023)	(0.023)	(0.024)	(0.030)	(0.086)
mgr_exp2	-0.001**	-0.001*	-0.001**	-0.001*	-0.001	-0.007**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
competitors_2	-0.057	-0.375	-0.250	-0.387	-0.508	5.142***
	(0.319)	(0.420)	(0.395)	(0.428)	(0.496)	(0.730)
competitors_3	-0.504**	-0.660**	-0.562*	-0.675**	-0.927***	4.078***
	(0.203)	(0.304)	(0.294)	(0.312)	(0.355)	(0.308)
competitors_4	-0.871***	-0.891***	-0.805***	-0.904***	-1.162***	3.655***
	(0.190)	(0.289)	(0.278)	(0.297)	(0.348)	(0.265)
orders_written	0.006***	0.007***	0.007***	0.007***	0.008***	-0.001
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
mat_paid_after_delivery	0.007***	0.004	0.006*	0.005	0.004	-0.001
	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)
sales_paid_before_delivery	0.004***	0.006***	0.005**	0.006***	0.007***	0.006**
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
subcontract	-0.104	-0.232	-0.185	-0.225	-0.255	-0.047
	(0.137)	(0.176)	(0.182)	(0.174)	(0.191)	(0.271)
reason_govt						-0.160
						(0.226)
Constant	-2.103***	-2.050***	-1.501**	-2.024***	-1.189*	-6.919***
	(0.309)	(0.592)	(0.643)	(0.577)	(0.714)	(0.756)
Interviewer Dummies	NO	YES	NO	YES	YES	NO
P-Value		0.0000		0.0163	0.9037	
Supervisor Dummies	NO	NO	YES	YES	YES	NO
P-Value			0.0000	0.4004	0.9670	
Interviewer-Supervisor Dummies	NO	NO	NO	NO	YES	NO
P-Value					0.9481	
Pseudo R-squared	0.149	0.266	0.274	0.274	0.228	0.141
Observations	1696	927	1083	927	661	564
Log-Likelihood	-580.5	-356.9	-378.4	-352.6	-314.0	-108.0
Chi-Squared	175.7	211.3	218.5	221.0	191.3	498.9
P Value	0.000	0.000	0.000	0.000	0.000	0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$

Dependent variable is a dummy=1 if firm is reticent, 0 otherwise. Heteroskedastic robust standard errors in parenthesis. Interviewer and supervisor dummies are only available for the second wave of the survey, so "*wave2*" is dropped from models 2 to 6.

5.5.5 Controlling For Misclassification Error Using The HAS-Probit Model

This study now turns to the potential problem of misclassifying some of the reticent managers as possibly candid. Using the methodology explained in Section 5.4.1 and Equation 5.4.1.3 estimates are derived for the coefficients on the variables which are believed to affect reticence and also for the misclassification errors: α_0 and α_1. Results are shown in Table 5.6. Table 5.6 shows estimates from the ordinary probit estimation and the misclassification-adjusted probit estimates (labelled HAS-Probit). Table 5.6 also shows the marginal effects from both models.

The ordinary probit model used in Table 5.6 is model 9 from Table 5.3. Coefficients for the HAS-Probit model are similar to those for the probit model. Imposing the restriction that the probability of observing a false positive is zero, the estimate of the probability of a false negative is 0.215. This is taken to mean that possibly 21.5% of the sample were misclassified into the possibly candid group when they were in fact reticent. This value is consistent with the result, from the previous chapter, that the proportion of reticent firms must be at least 16.9% (Table 4.9 in Section 4.5.3). Furthermore, using this result suggests that the level of guilt required to give rise to the observed distribution for the number of yeses must be 6.0%[96].

As expected, the absolute value of the marginal effects from the HAS-Probit model are larger that those of the ordinary probit model. Results from the HAS-Probit model suggest that a 10 percentage point increase in the percentage of written orders is associated with an increased probability of being reticent of 0.01 probability points. This also applies to the percentage of material paid for after delivery and the percentage of sales paid for before delivery. The coefficients on these three variables have the same sign, the same sized marginal effect, and are all significant at (at least) the 5% level.

The largest marginal effect from the HAS-Probit model came from the dummy variable for having more than 5 close competitors. A company with more than five perceived close competitors is 25 percentage points less likely to be reticent compared to a company with no perceived close competitors.

[96] $r = 0.215$ and $Pr(y_i^{observed}) = (0.5 + 0.5b)(1 - r) = 0.416)$. This suggests that $b = 0.05987$

Table 5.6: HAS-Probit Estimations For Reticence

	(9)	(HAS-Probit)	(9 - Marginal Effects)	(HAS - Probit - Marginal Effects)
secondary (d)	0.152*	0.179	0.028*	0.036
	(0.088)	(0.112)	(0.016)	
tertiary (d)	-0.164	-0.170	-0.027	-0.034
	(0.276)	(0.288)	(0.040)	
south (d)	0.321***	0.385**	0.060***	0.076**
	(0.086)	(0.166)	(0.017)	
wave2 (d)	0.851***	0.937***	0.130***	0.166***
	(0.112)	(0.206)	(0.014)	
mgr_experience	0.053***	0.059**	0.010***	0.013**
	(0.020)	(0.026)	(0.004)	
mgr_exp2	-0.001**	-0.001*	-0.000**	-0.000*
	(0.001)	(0.001)	(0.000)	
competitors_2 (d)	-0.028	0.003	-0.005	-0.006
	(0.319)	(0.406)	(0.055)	
competitors_3 (d)	-0.476**	-0.534*	-0.071***	-0.090*
	(0.202)	(0.282)	(0.025)	
competitors_4 (d)	-0.844***	-0.947***	-0.198***	-0.252***
	(0.189)	(0.317)	(0.054)	
orders_written	0.006***	0.006***	0.001***	0.001***
	(0.001)	(0.002)	(0.000)	
mat_paid_after_delivery	0.007***	0.008**	0.001***	0.001**
	(0.002)	(0.004)	(0.000)	
sales_paid_before_delivery	0.004***	0.005**	0.001***	0.001**
	(0.001)	(0.002)	(0.000)	
subcontract (d)	-0.100	-0.114	-0.017	0.022
	(0.137)	(0.155)	(0.022)	
Constant	-2.124***	-2.074***		
	(0.308)	(0.382)		
Other control variables	YES	YES		
$alpha_0$		0.000		
		(Imposed)		
$alpha_1$		0.215		
		(0.336)		
Pseudo R-squared	0.148			
Observations	1704	1704		
Log-Likelihood	-582.2	-582.1		
Chi-Squared	174.7	30.4		
P Value	0.000	0.033		

$*p < 0.10, **p < 0.05, ***p < 0.01$

Dependent variable is a dummy=1 if firm is reticent, 0 otherwise. Heteroskedastic robust standard errors in parenthesis. Other control variables include: gender of owner (dummy); age of owner (category); and size of firm (category). The coefficients on the other control variables are not statistically significant at the 10% level.

(d) for dummy variable.

5.6 Conclusion

The aim of this study was to investigate the factors that drive reticence amongst companies when answering randomised response questions. The characteristics of the managers and their dealings were explored as potential indicators of the reasons for reticence. The latent factors behind these characteristics include: social desirability bias; cognitive problems; and the effect of attitudes on responses. Reasons were provided to rule out the impact of cognitive problems and attitudes on reticence. The impact of trust and the risk of detection, respectively, were examined in detail. These two factors are thought to have an effect on reticence via a desire, by firm managers, to appear socially just.

Trust was measured using information on the nature of contracts for the procurement of goods/services and the sale of good/services. A higher: proportion of procured goods that were paid for after delivery, instead of before delivery or on delivery; proportion of sales where the payment was received before delivery, instead of on delivery or after delivery; and proportion of purchase orders that were written, instead of oral with a witness or oral without a witness, were interpreted to indicate a lower level of trust. The nature of the purchase or sales orders was not interpreted as a causal variable but an indirect measure of trust.

Evidence was found for both trust and perceived detection to influence reticence. This seems to suggest that companies are more willing to exchange potentially sensitive information with another agent if they have trust in that agent and if there is relatively little chance of them being identified. In the context of corruption, this suggests that people are more likely to admit to something if there is less of a chance that the information they pass on can/will be used against them.

The study gave evidence to rule out the possibility of interviewer and supervisor effects biasing responses and inducing reticence in the interviewees. Controlling for interviewer effects; supervisor effects; and interviewer-supervisor effects does not significantly affect the results. Findings were also able to rule out the possibility of political connections biasing the results. However, no significant relationship between political connectedness and reticence was found.

All questions were asked in the same order for all managers and all questions were worded in the same manner. Taken in full, these findings suggest that despite the absence of interviewer and supervisor effects, a lack of trust that the data will not be used against them might cause interviewees to be reticent when asked sensitive questions.

Given the discussion of the RR technique in this chapter and the previous chapter. The following chapter extends the analysis of Chapter 3 by looking at the impact of ethnic networks on the propensity to bribe.

6 Ethnic Networks And Corruption: Firm Level Evidence

Abstract

This study investigates whether the ethnicity of firm managers has any impact on whether or not they pay a bribe to local government officials. Micro-level data is taken from an individual-level and a firm-level survey in Nigeria; primary data concerning the firm managers' ethnic groups is also used. Indices of ethnic diversity are constructed and the ethnic group of firm managers is compared to that of the local government politician. Findings reveal that firm managers who belong to the same ethnic group as the local politician are less likely to pay a bribe than firm managers who do not belong to the same ethnic group as their local politician. This result is robust to the inclusion of standard variables which are believed to have an effect on the propensity to pay a bribe. Conditional on covariates, the propensity to pay a bribe is significantly higher in the South-Western region, particularly in Lagos. However, within Lagos there ceases to be a significant effect of ethnicity on the propensity to pay a bribe. Further results suggest that ethnicity is only important in regions with a relatively low index of ethnic diversity. In sufficiently heterogeneous regions (like Lagos) ethnicity ceases to have a significant impact on the propensity to pay a bribe. This suggests that ethnic discrimination in the payment of bribes is more pronounced in less ethnically diverse regions. The data seems to confirm this idea when ethnic diversity (measured by the index of ethno-linguistic fractionalisation) is measured at the local government level.

6.1 Introduction

Given the justification of the indirect questioning method (Chapter 4) and the investigation into the driving forces behind reticence in answering sensitive RR questions (Chapter 5), the current chapter extends the analysis of Chapter 3 by seeing what happens when ethnic networks are added to the list of potential influencers of corruption. This chapter proceeds by using primary collected data on firm managers' reports of their ethnic background; state of origin; and languages spoken. This information is used to generate national; state-level; and local government level measures of ethnolinguistic fractionalisation (ELF). The study also measures whether or not the manager belongs to the same ethnic group as the local government chairperson; and whether or not they speak the same language as the local government chair. This information is used in models similar to those in Chapter 3 to see the effect of ethnic diversity on the propensity to bribe.

The labour economics literature has empirically examined the effect of a teacher and pupil sharing the same ethnicity on pupil outcomes such as: exam marks, absenteeism, and subjective evaluations. Findings have shown that teachers tend to give higher subjective evaluations to students of the same ethnic group [Ehrenberg, Goldhaber & Brewer , 1995]; and students have a lower rate of absenteeism when their ethnic group matches that of the teachers [Farkas, Grobe, Sheehan & Shuan , 1990]. Evidence has been shown for a positive effect of a shared ethnicity on exam marks for some ethnic groups, but a zero effect for others [Klein, Le & Hamilton , 2001]. Nevertheless, these relationships are not restricted to the labour economics literature and there is a much wider base that the current study builds upon. For example [Bandiera, Barankay & Rasul , 2005] use a collection action framework and present results which suggest that, conditional on the share of workers that are of the same nationality, productivity amongst workers increases with the number of workers of the same nationality. A similar effect can also be found at the sub-national level where the characteristic that varies is ethnicity (See Section 6.2.1).

The current study looks at the relationship between ethnic background and economic outcomes in meetings between firm managers and local politicians. The proposition that discrimination based on nationality exists within the economy is not necessarily a controversial one within the literature. Studies have shown evidence for differing levels of pay; and for varying rates of assimilation into a labour market for foreign versus domestic workers; between migrants from different countries; and for migrants of varying levels of religious closeness, respectively [Dickens & McKnight , 2008, Pierné , 2013]. Part of this effect seems to come from the majority/minority ethnic group dichotomy, with agents in the minority groups being at more risk of discrimination [Pham, & Reilly , 2009, Baulch, Pham & Reilly , 2012]. Furthermore, evidence has also been found for wage inequality between people of different sounding accents within the same ethnic group [Grogger , 2011], suggesting that racial discrimination has not only a visual aspect but also an auditory one that can be based on speech syntax; morphology; acoustics; and/or phonology.

A related topic to ethnic discrimination is ethnic diversity. Some of the potential costs and the benefits of ethnic heterogeneity have been assessed in the literature [Alesina & La Ferrara , 2005, Moradi & Baten , 2005]. The level of ethnic diversity within a country can have policy implications for an economy. Given a fixed proportion of an ethnic group within an economy, a higher probability of being randomly paired with somebody from another ethnic group seems to be related with lower levels of nationalist sentiment among minority ethnic groups Masella [2013]; also, being exposed to a language at school that is different to one's mother-tongue seems to have a similar effect [Clots-Figueras & Masella , 2013]. The effect of personal connections on within-firm outcomes have also been investigated [Bandiera, Barankay & Rasul , 2009, 2010]; the present study builds on this work by looking at the effect of political connections on outcomes external to the firm.

The current paper applies a similar analysis to the study of corruption. It tests the effect of a shared ethnicity between a local politician and a firm manager on the propensity for corruption, in the form of the manager paying a bribe to the politician, to take place. It is crucial to note that this study focuses on bribery and not other forms of corruption that might be affected by the existence of a shared ethnicity between both parties.

This paper is the first empirical study to look at the links between ethnicity and corruption at the micro-level. Using a set of unique data sources this study identifies the ethnic identities of household individuals; firm managers; and local government politicians. This data is used to provide evidence for the existence of ethnic networks in business and the effect of such networks on the payment of bribes to public officials.

The study is the first of its kind to use disaggregated data that is able to identify the ethnic group of a person. The data used in this investigation covers the individual; the household; the local government area; and the geo-political state. The primary research question is whether a firm manager who belongs to the same ethnic group as a local politician will be more or less likely to pay a bribe to that politician in the running of his/her business operations. The study identifies two plausible channels through which a shared ethnicity might affect the propensity to bribe: redistribution along ethnic lines will lead to a positive effect of a shared ethnic group on bribery as similar ethnic groups are more willing to pay their kin a bribe rather than a non-kin member; discrimination along ethnic lines will lead to a negative effect of a shared ethnic group on bribery as dissimilar ethnic groups are extorted for more bribes.

The study also uses information on the languages spoken by the managers to see the impact that this might have on the propensity to bribe. Speaking the same language as the local politician might allow some non-local managers to gain rapport with the politicians; or pretend that they are of the same ethnic group. This might have an impact on their propensity to bribe the politician.

The secondary research question involves the use of an index of ethnolinguistic fractionalisation (ELF). Combined with the information regarding a similar ethnicity this paper investigates whether the effect of having the same ethnicity as the local politician on bribery varies depending on the level of ELF in the region (state and local government area, respectively). The study also considers a potential non-linear effect of ELF on bribery. With the results of the aforementioned analysis, the study uses ethnic networks as an instrument for bribery to see if bribe paying firms are more or less likely to to report being the victim of a crime (See Appendix A.4.3).

Thus, the study addresses research questions at the firm level; local government level; and state level. The study looks at the effect of a shared ethnicity on bribery at the firm level; while the effect of ELF on corruption is investigated at the local government level and the state level. It is important to look at the relationship between sub-national ELF and sub-national corruption because many of the claims of a relationship between ELF and corruption are based of cross-national data despite being used to make inferences about sub-national economic activity. The current investigation intends to shed light on the story behind ELF and corruption and what might be driving the relationship.

Results from the analysis suggest that ethnic networks matter in the payment of bribery: belonging to the same ethnic group as the local politician is associated with a reduced probability of paying a bribe. Thus, the effect of discrimination seems to outweigh the effect of kinship redistribution. This suggests that bribery can occur through (discriminating) ethnic networks. Secondly, the effect of ethnic networks is most pronounced in areas with low levels of ethnic diversity. In cosmopolitan regions, where people are more used to dealing with different ethnicities, the effect of sharing the same ethnicity as the local politician disappears. This suggests that discrimination is more prevalent in less diverse regions and bribe discrimination will be reduced the more cosmopolitan a society is. Finally, amongst firms that have been a victim of crime, bribe paying firms are less likely to report the crime. This suggests that estimates of business malpractice will be reduced in corrupt environments.

Thus, this study makes the following major contributions: it uses disaggregated individual level; household level; firm level; local government level; and state level information on the ethno-linguistic makeup of Nigeria. Despite the absence of information on ethnicities from the national census and government data sources, this study is able to generate estimates of local government level and state level ethnolinguistic fractionalisation. These indices of ELF are calculated using over 90 ethno-linguistic groups. The study generates a method for the analysis ethnic networks; this is labelled the Similar-Ethnicity variable. The study also applies the analysis of the relationship between ELF and corruption to the micro level, using the Similar-Ethnicity variable gives significant results which have not been done before using empirical data.

6.2 Literature Review

This section reviews the literature relating to bribery and ethnicity in the field of economics. Recent evidence for a link between ethnicity and patronage has been provided by Burgess, Jedwab, Miguel, Morjaria & i Miquel [2011] and Do, Nguyen & Tran [2013]. Both of these studies tell a story about individuals in power passing on resources to members of their own ethnic group instead of the population as a whole. Whilst some evidence exists for politicians allocating resources to other ethnic groups, this seems to be driven by a need to get more votes in order to win an election, rather than altruism or a lack of patronage on ethnic grounds. Burgess, Jedwab, Miguel, Morjaria & i Miquel [2011] focus on Kenya and document the construction of roads in areas dominated by people of the same ethnicity as the political leader whilst Do, Nguyen & Tran [2013] talks about the distribution of resources, by recently promoted officials in Vietnam, to their home districts. The flipside of this is that some individuals might extort people from other ethnic groups while choosing not to do so to members of their own ethnic group. The current study considers both of these explanations while looking at the supply of bribes in Nigeria. It could be the case that ethnicity drives firm managers to be more willing to transfer resources to public officials of the same ethnic group; on the other hand it might be the case that ethnicity drives public officials to demand more bribes from firm managers of a different ethnic group.

6.2.1 Some Concepts In The Literature On Ethnicity

This section considers three respective ways that ethnic diversity can directly affect economic choices, these are: through preferences; strategies; and the production function. Results from the social identity literature suggest that individuals might derive utility from the well-being of members of their own ethnic group and disutility from the well being of members of other ethnic groups [Tajfel, Billig, Bundy & Flament , 1971]. Alesina & La Ferrara [2000] formalise this concept by analysing group participation within a heterogenous population, where the utility gained from joining a group is positively related to the proportion of groups members who have the same type as oneself; and negatively related to the proportion of other types.

Ethnic diversity can affect economic activity by influencing the strategies that individuals play. In the absence of any preferences for or against heterogeneity it might be optimal to give preferential treatment to one's own ethnic group. This might occur due to market imperfections: e.g. to maintain a reputation mechanism by exchanging information on opportunistic behaviour in the presence of asymmetric information [Greif , 1993]. However, preferential treatment of ones own ethnic group can occur in the absence of market imperfections. When contracts cannot be legally enforced, being a member of an ethnic group increases the set of dominating group strategies, this is because reciprocity and punishment can be more direct at both the individual and the other members of the group [La Ferrara , 2003].

Finally, in a production function where more variety in intermediate inputs leads to increased output, one can interpret a higher variety of individual skills as having the same effect [Alesina, Spolaore & Wacziarg , 2000]. In this example the costs of heterogeneity lie outside the production function, as a result, more heterogeneity is always better than less when trying to maximise output. When incorporating the costs of heterogeneity, for example the difficulty in communication between people who do not speak the same language or have different methods, then there is an optimal level of heterogeneity which will lie below that of the previous example. This optimal point will depend on the trade-off between the gains from heterogeneity and the costs of heterogeneity [Lazear , 1999a,b].

Recent work on ethnic groups has posited that higher levels of ethnic heterogeneity is associated with more social cohesion [Bécares, Stafford, Laurence & Nazroo , 2011]. Other studies have highlighted that an increased ethnic density (defined as the proportion of ethnic minorities in a region) is associated with: lower levels of discrimination; a smaller negative effect of discrimination on outcomes; and improved health outcomes [Bécares, Nazroo & Stafford , 2009]. If this is the case, then it might follow that increased diversity will lead to a lower level of discrimination based on ethnic lines. The arguments in the previous two paragraphs support the idea that the more heterogeneous a population, then the less of an impact that ethnicity has on economic outcomes. I.e. ethnicity only matters if there is an obvious majority or minority; the more fractionalisaed an area is then the less likely it is that ethnicity has an impact on economic outcomes. This implies that more ethnically homogeneous regions will see a larger effect of racial discrimination or racial patronage than lesser ethnically homogeneous regions. This suggests a negative relationship between corruption and ethnic fractionalisation. Cerqueti, Coppier & Piga [2012] argue for an inverse-U shaped relationship between corruption and fractionalisation; the highest levels of corruption are seen at the lowest and highest levels of fractionalisation. The previous argument is not at odds with this result so long as the data for Nigeria lies on the downward sloping section of the theoretical corruption-fractionalisation curve.

6.2.2 Shortcomings Of The Literature And The Contribution Of This Chapter

When trying to see if politicians favour members of their own ethnic group at least two things seem to be of importance: the ethnic group of the favoured/unfavoured individual; and the ethnic group of the politician doing the discrimination. Assigning ehnic groups to politicians and other individuals requires individual level data concerning both agents, however, the literature on ethnicity and economic outcomes tends to aggregate the ethnicities of a group of people living in the same geographic location [Burgess, Jedwab, Miguel, Morjaria & i Miquel , 2011, Do, Nguyen & Tran , 2013, Michalopoulos & Papaioannou , 2013]. Some studies base their categorisation of ethnicities on ethnographic atlases that were created in the late 1960 but which have been dismissed by anthropologists since their creation [Moradi, 2013][97]. Data based on historical locations of ethnic groups might be wrong due to subsequent migration of individuals from those groups to other areas.

Other studies use the majority ethnic group as a measure of the likely ethnicity of the politician [Isaksson , 2013]. Doing this makes the results subject to measurement error because it is possible for minority ethnic politicians to be elected to political positions; and it is possible for minority ethnic groups to be more active in a business environment. Such criticisms should be taken seriously given that the share of the largest ethnic group, in some cases, is less than 0.2; e.g. Tanzania and Kenya [Ibid.]. In cases such as Nigeria, where the share of the largest ethnic group is higher (0.33), there exist other ethnic groups that maintain a considerable size (0.21 and 0.22) which make it likely for a politician to hail from one of these ethnic groups. Moreover, using the largest ethnic group as a proxy for the ethnicity of the politician seems to be in contradiction with the literature that uses geographic location as a proxy for ethnicity. If the majority group lives in one part of the country and the other groups live in other parts then being of a minority group does not necessarily mean that the local politician will not be a co-ethnic.

Furthermore, when considering the issue of corruption along ethnic lines, an assumption that a majority ethnic group will only vote in a member of the same ethnic group seems to imply that, upon election, the official will show a preference to his/her own (majority) group. This nullifies the need for a study on favouritism along ethnic lines because it begs the question. Using individual-level data on

[97] Correspondence with Author.

ethnicities allows one to see the effect of within-region variation in ethnicity on the occurrence of corruption. Studies that use regional aggregated data concerning ethnic groups only allow for a between-region analysis of ethnicity and corruption.

Other work [Isaksson , 2013] uses the home language spoken by an individual as a proxy for their ethnic group. The current study goes one step further by using data on all languages spoken by the individual; and the self-reported ethnicity of the individual. Finally, while capturing aspects of diversity, indices of fractionalisation and polarisation do not measure individual level differences between politicians and the populace [Alesina, Devleeschauwer, Easterly, Kurlat & Wacziarg , 2003, Montalvo & Reynal-Querol , 2005b, Alesina & La Ferrara , 2005]. The current study bypasses all of these problems because it measures ethnic group via self-reported survey responses. The survey used in the current study also asks the languages spoken by each individual.

6.3 Data And Variables

This section describes the data used in this study and the construction of variables. The data used in the analysis comes from three main sources. One of the main sources for the firm level data is the Nigerian Bureau Of Statistics (NBS) and the Economic And Financial Crimes Commission (EFCC) Business Survey On Crime And Corruption And Awareness Of EFCC In Nigeria (2007). Individual level data is taken from the Centre For Law Enforcement Education (CLEEN) 2010 National Crime And Safety Survey. These datasets were augmented by retrieving primary data about the ethnicity, religion, gender, state of origin, and lingualism of the business managers. The NBS/EFCC survey contains the information about bribery; the CLEEN survey contains information that is used to create the indices of ethnolinguistic fractionalisation (ELF) within Nigeria at the state level; and local government level. The primary data is also used to create indices of state and local government level ELF within Nigeria.

6.3.1 Information On Individuals For The Construction Of Individual-Level Index Of Ethnic Fractionalisatioon

To the best of the author's knowledge, no data exists for the current level of ethnic diversity in Nigeria. Most of the current sources of information are either unobserved[98]; old [Taylor & Hudson , 1972]; or uncomprehensive [Central Intelligence Agency , 2013]. These sources do not represent a necessarily accurate view of the 250+ ethnic groups contained within the nation . This lack of recent data in ethnicities poses a potential problem for the analysis of ethnicity and bribery. In

[98] Correspondance with the National Population Commission (2010).

order to solve the problem of lack of data on ethnicities in Nigeria this study uses information from a representative household survey that was conducted in all 36 geo-political states and the Federal Capital Territory. The survey was conducted by the CLEEN Foundation between February and May 2010. Information on this data is given in Section 2.5.

6.3.2 Information On Company Characteristics (Including Bribes Paid)

The firm-level survey contains a set of questions relating to bribery: whether or not the firm has paid a bribe; how much was paid and what the bribe was paid for. Different types of bribe are measured: bribes during police investigations; bribes to avoid traffic offences; bribes to customs officials to clear goods faster; and bribes to the court, were some of the different measures that are included in the data.

The survey also included the street address of each company. This study matches these addresses to their respective local government area using LG information from the Independent National Electoral Commission [INEC , 2010] and the City Population database [Brinkhoff , 2013]. The INEC [2010] database contains information on all the political wards within Nigeria, matched to their respective local government area, which in turn are matched to their respective geo-political state. The Brinkhoff [2013] database contains a geographical mapping of similar information; it shows the physical location of each LGA within each state in Nigeria.

Information concerning whether or not the company was a victim of crime; and whether or not the company reported the crime is also used. Companies were asked whether they had been a victim of theft; vandalism; theft of vehicles; theft from vehicles; robbery; assault; theft/fraud by employees; or theft/fraud by outside stakeholders in the past year, respectively. Regarding each crime, they were also asked if they reported the act to the police. The current study uses this data by constructing a dummy indicating whether or not the firm was a victim of at least one of the crimes mentioned; and another dummy indicating whether or not at least one crime went unreported.

6.3.3 Information On Company Managers - Ethnicity; State Of Origin; Languages Spoken; Religion; And Gender

The primary data on the ethnic group; gender; religion; languages spoken; and state of origin of the managers of the respective companies in the NBS/EFCC survey is used in this study. This data was collected by the author, by contacting the companies directly. Telephone calls were made between November 2012 to December 2012 to each of the 2,110 businesses in the sample. Out of the 2,110 companies, 1,267 companies were successfully contacted, this represents 60% of the total sample. Results from all 1,267 managers were analysable. Since the NBS/EFCC survey was conducted in 2007, information was collected about the managers in 2007; and not the current manager.

Information about the ethnicity and state of origin of the managers allows this study to account for the movement of labour within the borders of Nigeria, it also removes the restriction of assigning one ethnic group to a geographic area. This study combines the data on ethnicity with the information on the LGA of the company to create indices of ELF at the state; and local government levels as well as measures of ethnic networks at the local government level.

As previously discussed, the current study benefits from the fact that it uses micro data on the ethnicities of business managers and local government politicians. This data is useful because it allows one to test for whether a match between the ethnicity of the manager and the politician has any impact on the propensity to bribe, and the size of bribe payment. The previous literature (discussed in Section 6.2) makes use of aggregate data on ethnic diversity, corruption and economic outcomes. Whilst potentially useful, aggregate studies fail to take into account the unobserved heterogeneity at the firm level which might not show up at the regional level. The data on ethnicity allows this study to investigate this issue which has not been a feature of the previous literature.

A common ethnicity between the business manager and local politician might reduce the propensity the bribe, increase the propensity to bribe, or have no effect on the propensity to bribe. In the case of discrimination against other ethnicities, the match would be associated with a reduced probability of bribing. In the case of revenue sharing [Angelucci, de Giorgi & Rasul , 2012] (where a manager is more willing to pay a bribe to his or her own ethnic group) the match would lead to an increased propensity to bribe. There might be other reasons for the effect to be pushed one way or another; for example politicians from some ethnic groups might be better at extracting bribes from managers of other ethnic groups; another potential reason is that managers from some ethnic groups might be less willing to part with their money (in order to pay a bribe) than managers from other groups.

The main variable of interest in this study is an indicator that measures whether or not the ethnic group of the firm manager is the same as the ethnic group of the local government politician. This variable, labelled "Similar Ethnicity", is a dummy variable equal to one if the ethnicities are the same, 0 otherwise. The ethnicities of the local politicians were acquired by using their names to derive their ethnic origin [Nicoll, Bassett & Ulijaszek , 1986, Coldman, Braun & Gallagher , 1988, Angelucci, de Giorgi, Rangel & Rasul , 2010, Angelucci, de Giorgi & Rasul , 2012].

The firm level dataset consists of 61 ethnic groups and the individual level dataset has 99 groups. In order to make the information more useable; and reduce the number of potentially empty cells, this study categorises the ethnicities into 4 main groups. These are the 3 most populous ethnic groups in the country: Hausa; Yoruba; and Ibo, and another group, Other, that captures all other ethnic groups contained in the dataset. Covariates measuring factors influencing the propensity to pay a bribe based on Chapter 3 [Svensson , 2003] are also included in the analysis.

6.4 Methodology

This section describes the methodology used in the subsequent analysis. Similar-Ethnicity is a dummy variable equal to 1 if the ethnicity of the firm manager, as defined in Section 6.3.3, is the same as the ethnicity of the local government chairperson, 0 otherwise. This is shown in Table 6.1.

Table 6.1: Similar Ethnicity

		Ethnicity Of Local Politician			
		Yoruba	Igbo	Hausa	Other
Ethnicity Of Firm Manager	Yoruba	1	0	0	0
	Igbo	0	1	0	0
	Hausa	0	0	1	0
	Other	0	0	0	1

ELF is defined as the probability that two randomly selected individuals from a population will not belong to the same ethnolinguistic group. Considering a country that has a population of N individuals who are distinguished by K ethnolinguistic groups, representing each group by $i = 1, ..., K$. Each individual belongs to one group only and the number of people in group i is represented by N_i. The current study allows for flexibility by not imposing any restrictions on the geographical location of any member of a group, therefore, individuals of ethnolinguistic group i can all live in the same region or be seperated and live in different regions. Since each individual belongs to only one group the sum of group populations will be equal to the size of the total population: $N = \sum_{i=1}^{K} N_i$. The share of the population belonging to ethnolinguistic group i can be expressed as the ratio of the size of ethnic group to the total population: $s_i = \frac{N_i}{N}$. Therefore the sum of population shares will equal 1: $\sum_{i=1}^{K} s_i = 1$

ELF is defined as the probability that two randomly selected individuals will not belong to the same group. Formally:

$$
\begin{aligned}
ELF =& 1 - \sum_{i=1}^{K} s_i^2 \\
=& \sum_{i=1}^{K} s_i - \sum_{i=1}^{K} s_i^2 \\
=& \sum_{i=1}^{K} s_i - s_i^2 \\
=& \sum_{i=1}^{K} s_i \left(1 - s_i\right)
\end{aligned} \tag{6.4.0.1}
$$

The index of ELF satisfies the primary requirements of a diversity measure as stated by Shannon [1948]:

- For a fixed number of groups, the measure reaches a maximum when all group sizes are equal
- If all groups sizes are equal, then the index increases with the number of groups in a society

Using the notation of Chapter 3, the following model[99] is estimated:

$$
Pr[Bribe \quad Dummy_i = 1] = \Phi(\beta_1 Similar_i + \beta_2 ELF_l + \beta_3 (Similar_i \times ELF_l) + \beta_C^{'} C_i) \tag{6.4.0.2}
$$

Where $Bribe \quad Dummy_i$ is a dummy equal to one if the firm paid a bribe, 0 otherwise; $Similar_i$ is the value of Similar-Ethnicity for firm i; ELF_l is the index of ethnolinguistic fractionalisation for local government area l; $(Similar_i \times ELF_l)$ is the Similar-Ethnicity-ELF interaction term; C_i is a vector of explanatory variables denoting control rights of public officials over firms; and Φ is the standard normal distribution function.

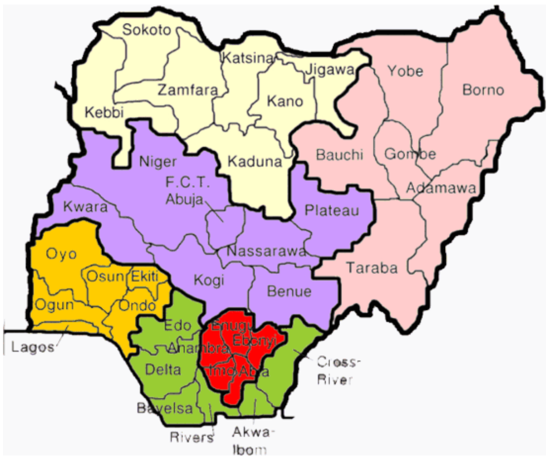
6.5 Results

This section presents and interprets the results from the data. Additional tables and figures are located in Appendix A.4. Definitions and descriptions of the main variables are included in Table 6.2. Summary statistics are shown in Tables 6.3 to

[99] The relationship between corruption and ELF, at the state and local government levels, respectively, is investigated the Appendix A.4 Section A.4.3. At these levels, the following model is estimated: $Corruption_s = \beta_0 + \beta_1 ELF_s + \beta_2 ELF_s^2 + \beta_3 X_s' + \xi_s$.

6.5. The results from the main set of estimations are displayed in Tables 6.6 to 6.8. A map of Nigeria showing the different geo-political zones and states is shown in Figure 15. A map of Nigeria showing the ethnic composition of the country using firm level data and individual level data, respectively, is shown in Figure 16. All probit estimations use Huber-White heteroskedasticity robust standard errors unless otherwise stated. The sample size for the entire country consists of 2,110 firm managers. The sample size for the individual level data, used to compute the ethnic composition of the country and regional indices of ethnic fractionalisation, is 10,228. All ethnicities were treated as seperate when constructing the Similar-Ethnicity variable. This is in contrast to Table 6.1 and Figure 16 which combine all ethnic groups that are neither Hausa, Igbo nor Yoruba into the group: "Other".

Figure 15: A Map Of Nigeria Indicating The Different Geo-Political Zones (NC,NE,NW,SE,SS,SW)



Source: Nairaland.com

Summary statistics for the firm-level variables are shown in Table 6.3. This table shows data concerning the 1,267 firms which have information on ethnicity; and the 843 firms which do not have information on ethnicity. There are no significant differences between the two groups in revealed propensity to import or export directly ("*Trade*") or in the usage of security services ("*Security*"). Nevertheless, there seem to be significant differences in the paying of bribes; number of employees; degree of foreign ownership; and age of the firms. On average, firms with data on ethnicity are 0.045 probability points more likely to pay bribes while firms without data on ethnicity have more employees; are 0.06 probability points more likely to be owned by a foreign individual or company; and are 4 years older than the others.

Concerning the 1,267 firms with information on ethnicity, 28.7% admitted to having paid a bribe in the year leading up to the survey; the median firm had less than 50 employees; 7% of the firms had engaged in the direct importing of exporting of goods and/or services in the previous year; 11% had foregin ownership which was equal to or exceeded 25% of the firm's shares; 99% spent money on some form of security in the previous year; the mean age among the 1,267 firms was slightly above 16 years; and 62.7% of the sample had firm managers with the same ethnicity as that of the local government chair ("Similar-Ethnicity").

Table 6.2: Data Definitions - Variables That Are Measured At The Firm Level

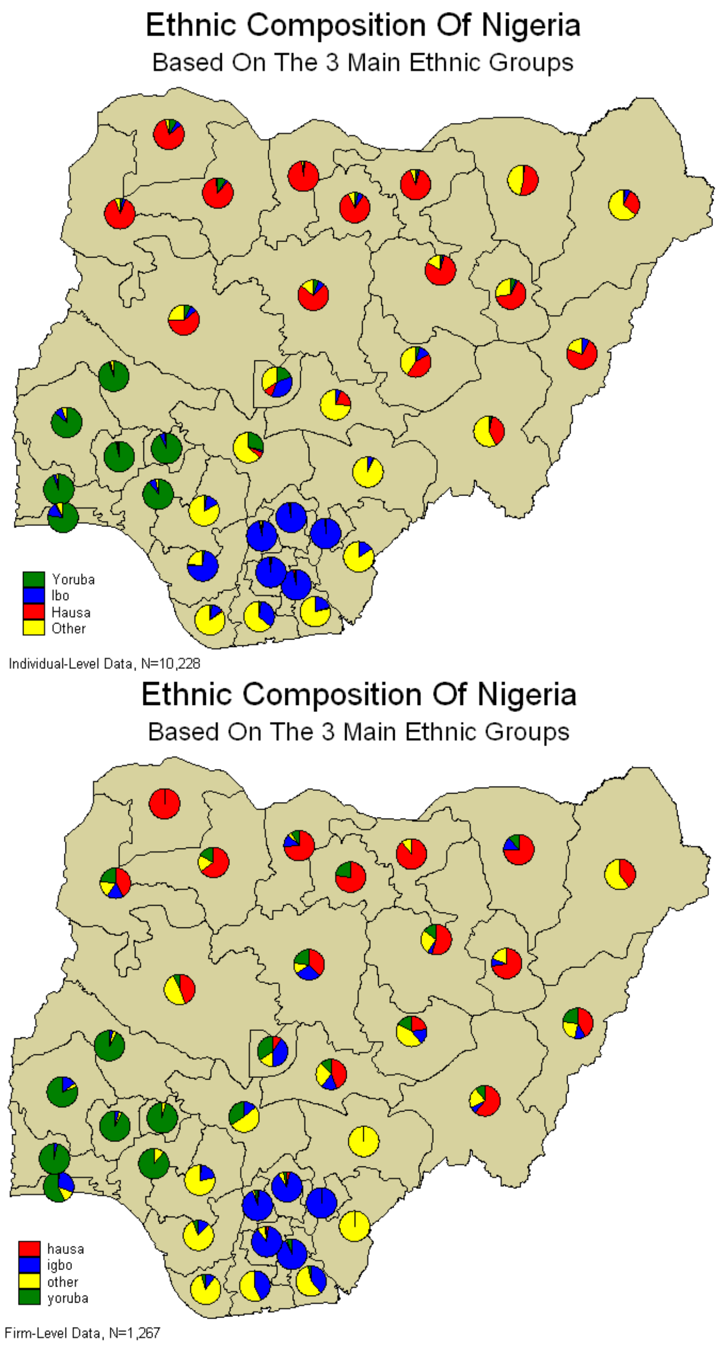
Category	Variable Name	Definition	Measurement
Dependent Variable	Bribe Dummy	Dummy=1 if firm admitted to bribing; 0 otherwise	{0;1}
Independent Variables	Similar Ethnicity	Dummy=1 if ethnicity of manager=ethnicity of local government chairperson, 0 otherwise	{0;1}
	Different Ethnicity	1-Similar Ethnicity	{0;1}
	Employee	number (category) of paid employees {Less Than 50; 50 to 100; 100 to 250; Over 250}	{1; 2; 3; 4}
	Trade	Dummy=1 if firm engages in international trade, 0 otherwise	{0;1}
	Foreign	Dummy=1 if foreign ownership $\geq$25%, 0 otherwise	{0;1}
	Security	Dummy=1 if company using security service, 0 otherwise	{0;1}
	Lage	Log of firm's age	

Table 6.3: Summary Statistics For Variables Measured At The Firm Level: Selected & Unselected Subsamples; And For Entire Sample

Variable	Statistic	Firms with Data on Ethnicity	Firms without Data on Ethnicity	All Firms	Test of Relationship (P and Z Values)
Bribe Dummy	Min	0	0	0	
	Max	1	1	1	
	Mean	.287	.242	.269	
	Median	0	0	0	0.022^{+}
	Std. Dev.	.45	.43	.27	
Employee	Min	1	1	1	
	Max	4	4	4	
	Mean	1.53	2.18	1.79	
	Median	1	2	1	0.000^{+}
	Std. Dev.	.95	1.29	1.15	
Trade	Min	0	0	0	
	Max	1	1	1	
	Mean	.07	.09	.08	
	Median	0	0	0	0.167^{+}
	Std. Dev.	.26	.28	.26	
Foreign	Min	0	0	0	
	Max	1	1	1	
	Mean	.11	.17	.13	
	Median	0	0	0	0.000^{+}
	Std. Dev.	.316	.372	.34	
Security	Min	0	0	0	
	Max	1	1	1	
	Mean	.99	.99	.99	
	Median	1	1	1	0.594^{+}
	Std. Dev.	.08	.09	.08	
Age	Min	0	0	0	
	Max	90	90	90	
	Mean	16.3	20.26	17.9	
	Median	14	17	15	0.000^{-}
	Std. Dev.	13.2	15.43	14.26	
Lage	Min	0	0	0	
	Max	4.50	4.50	4.50	
	Mean	2.45	2.68	2.54	
	Median	2.64	2.86	2.71	0.000^{-}
	Std. Dev.	.94	.94	.94	
Similar Ethnicity	Min	0			
	Max	1			
	Mean	.627			
	Median	1			
	Std. Dev	0.484			
N		1,267*	843*	2,110*	

In each column, for each variable, the table reports the minimum, maximum, mean, median and standard deviation for each (sub)sample. The last row details the number of observations for each (sub)sample. *Only 1255, 834, and 2,089 observations for "Lage", respectively, due to log(0) being undefined. $^{+}$ P-Value was computed using a chi-squared test of independence between the variable and a "selection" variable equal to 1 if the managers ethnicity is known, 0 otherwise. The test on "employee" has 3 degrees of freedom, the tests on "bribe dummy", "trade", "foreign" and "security" each have 1 degree of freedom. $^{-}$ Z-Value calculated using a Wilcoxon-Mann-Whitney (rank-sum) test.

Figure 16: Ethnic Composition Of Nigeria



The most dominant ethnicity amongst local politicians in this dataset is the Yoruba tribe whose politicians take up 37% of the local government chairs for the local governments of the 1,267 firms selected for the sample. This is followed by the Hausa tribe (28%); then the Igbo tribe (18%); and then other tribes (17%). Amongst firm managers, 32% are Yoruba; 29% are Igbo; 22% come from one of the other tribes; and 17% are Hausa. Thus, the Yoruba ethnic group make up the bulk of the managers and LG chairs in this dataset; while, compared to the Igbo ethnic group, the Hausa group are over-sampled among the set of LG chairs and under-sampled among the set of firm managers. Elements along the main diagonal of Table 6.4 constitute firms whose managers share the same ethnicity as the local politician, i.e. "Similar-Ethnicity" equals 1. Elements on the off diagonal constitute firms whose managers are of a different ethnic group than the local politician, so for these firms "Similar-Ethnicity" is equal to 0.

The majority of the cases where Similar-Ethnicity is equal to 1 (38%) come from firms in the South-West region of the country (Table 6.5), however, this region also contains the majority of the firms in the dataset (34%). In order, to see where the matches are most likely to come from the last column of Table 6.5 also shows the probability of observing a match conditional on being in the region: (Pr(Similar Ethnicity $= 1|$Zone $= x$)).

Table 6.4: Cross Tabulation Of The Ethnicities Of Firm Managers And Local Government Chiefs

		Ethnicity Of Local Government Chief				
		Yoruba	Igbo	Hausa	Other	Total
Ethnicity Of Firm Manager	Yoruba	331	6	55	11	403
	(%)	26.12	0.47	4.34	0.87	31.81
	Igbo	93	192	53	35	373
	(%)	7.34	15.15	4.18	2.76	29.44
	Hausa	4	3	185	18	210
	(%)	0.32	0.24	14.60	1.42	16.57
	Other	43	23	64	151	281
	(%)	3.39	1.82	5.05	11.92	22.18
	Total	471	224	357	215	1,267
	(%)	37.17	17.68	28.18	16.97	100.00

Pearson χ^2_9 statistic = 1300. P-value=0.000. Percentages represent the percentage of the sample of 1,267 firms that exist within the respective cell. The main diagonal represents firm managers whose ethnicities match those of the local government chief. The off diagonal entries are firm managers who have a different ethnicity to that of the local government chiefs when splitting ethnicities into the 4 main groups of: Yoruba; Igbo; Hausa; and Other.

Table 6.5: Cross-Tabulation Of Ethnic-Match And Zones

Geo-Political Zone	Similar Ethnicity Mismatch	Match	Total	$Pr(Match\|Zone)$
NC	120	61	181	0.33
	66.30	33.70	100.0	
NE	44	54	98	0.55
	44.90	55.10	100.0	
NW	64	113	177	0.64
	36.16	63.84	100.0	
SE	17	184	201	0.92
	8.46	91.54	100.0	
SS	93	78	171	0.46
	54.39	45.61	100.0	
SW	135	304	439	0.69
	30.75	69.25	100.0	
Total	473	794	1,267	0.63
	37.33	62.67	100.00	

Pearson χ^2_5 statistic = 168 P-value=0.000.

6.5.1 Effect Of Similar Ethnicities On Bribery

The summary statistics establish that managers and politicians, respectively, differ in their ethnicity; and that the probability of sharing the same ethnicity as the local politician also differs across regions and ethnicities. The primary question of interest is whether the latter phenomenon is correlated with bribery. Despite the fact that without exogenous variation in ethnic networks this study cannot identify the causal impact of ethnicity on bribery, establishing whether the two are correlated can be a powerful tool to test whether ethnicity captures meaningful differences across managers, instead of random noise; and to establish the practical importance of ethnic networks in the study of firm level bribery.

The first main results are shown in Table 6.6. In this table the dependent variable is a dummy equal to one if the firm admitted to paying a bribe, 0 otherwise. Similar-Ethnicity is included as an independent variable with other factors which influence the payment of bribes [Svensson , 2003]. Including the matching variable with the other independent variables does not significantly alter their sign or magnitude. Similar-Ethnicity enters negatively and significantly, as expected, suggesting that the power of discrimination against other ethnic groups is greater than that of income sharing among members of the same group.

Similar-Ethnicity enters significantly in models 1 to 4 of Table 6.6. Models 1 to 3 control for the zone in which the firm resides. Therefore, when controlling for zone, the probability of paying a bribe is significantly lower for managers who share the same ethnicity as the local politicians compared with managers who do not share the ethnicity of the local politician. Results from Table 6.7 show that the inclusion of the ethnicity variable does not significantly alter the sign or magnitude of the zonal dummies when trying the explain the propensity to bribe. All zonal dummies have negative coefficients and the omitted zone is the South-West region, indicating that, on average and ceteris paribus, firms located in the South-Western region of Nigeria are more likely to pay a bribe compared with firms located elsewhere within the country. Out of all firms in the sample that were located in the South-West region, 39% admitted to bribery; whereas the next highest figure was 29% which was for the North-Central region. The South-West zone is home to the commercial capital of Lagos, while the North-Central Zone is home to the Capital city of Abuja. These results add support for the idea that there is a higher propensity to bribe in areas with greater economic and political activity. Results from this table do not change when clustering the standard errors by zone or local government area.

Models 4 and 5 of Table 6.6 investigate this by including only 1 zonal dummy: for the South-West region (Model 4); and including a dummy variable for Lagos State along with the 5 original zonal dummies. In model 4 the South-West dummy has a positive and significant (1%) coefficient. Using the South-West dummy instead of the 5 other dummies does not alter the signs or statistical significance of any other coefficient. On the other hand, when the Lagos dummy is entered into the model along with the 5 original zonal dummies (model 5) the coefficient on Lagos is positive and significant at the 1% level but the coefficients on the 5 dummy variables are all positive and significant. The coefficient on Lagos is larger and more significant than the coefficients on the 5 zonal dummies. These results suggest that the negative coefficients on the zonal dummies in Models 1 to 3 were driven by Lagos State and not the other states in the SW region.

Another result from this Table is that larger firms (those with over 250 employees) are less likely to pay bribes, on average and ceteris paribus, than other firms. This result is consistent with Delavallade [2011] who suggests that smaller firms are more likely to engage in petty corruption than larger firms, who are more likely to engage in state capture. The coefficient on international trade is positive and statistically significant which hints at the role of public officials at ports (e.g. customs, police) in extracting bribes from firms.

When other firm characteristics (a dummy for foreign ownership; a dummy for use of a security service; and the log of age) are added to models 3, 4 and 5 (results omitted) of Table 6.6 the coefficient on Similar-Ethnicity changes by +0.008; +0.005; and +0.002, respectively. Hence, these coefficients are not significantly different from the ones presented in the table.

The general results concerning a shared ethnic group and ethnolinguistic fractionalisation can be seen from the results of Table 6.8. These models interact

Table 6.6: Firm Level Probit Estimations Of Bribery On Matching, With Zonal Dummies

Dependent Variable:	Dummy=1 if firm admitted to bribing; 0 otherwise				
	1	2	3	4	5
Similar Ethnicity	-0.194**	-0.164*	-0.147*	-0.141*	0.032
	(0.084)	(0.086)	(0.086)	(0.080)	(0.090)
Trade		1.057***	1.119***	1.127***	0.928***
		(0.149)	(0.152)	(0.150)	(0.165)
Employee==50-100			0.139	0.118	0.088
			(0.123)	(0.121)	(0.128)
Employee==100-250			-0.183	-0.185	-0.223
			(0.143)	(0.144)	(0.150)
Employee==Over 250			-0.378**	-0.375**	-0.382**
			(0.155)	(0.157)	(0.160)
Zone==SW				0.421***	
				(0.080)	
Region==Lagos					1.575***
					(0.168)
Constant	-0.148*	-0.283***	-0.267***	-0.690***	-1.483***
	(0.084)	(0.088)	(0.092)	(0.069)	(0.168)
Zonal Dummies	YES^{--}	YES^{--}	YES^{--}	NO	YES^{++}
Pseudo R-squared	0.031	0.068	0.074	0.069	0.143
Observations	1267	1267	1267	1267	1267
Chi-Squared	46.09	90.96	97.77	89.64	178.23
P-value	0.000	0.000	0.000	0.000	0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$. $^{++}(^{--})$: Plusses (minuses) indicate that the coefficients on all zonal dummies are positive (negative) and significant at, at least, the 5% level. Models 1-5: Heteroskedasticity Robust Standard Errors In Parenthesis. Similar Ethnicity: Dummy=1 if ethnicity of manager matches ethnicity of LGA Chair, 0 otherwise. Unit of observation is the firm. Similar Ethnicity is calculated using manager data collected by the Author. All other variables are calculated using data from the NBS/EFCC survey. "Zone==..." is a dummy equal to 1 if the firm is located in said geo-political zone, 0 otherwise. Zones are defined as: NC= {Benue, Kogi, Kwara, Nassarawa, Niger, Plateau, FCT (Abuja)}; NE={Adamawa, Bauchi, Borno, Gombe, Taraba, Yobe}; NW={Jigawa, Kaduna, Kano, Katsina, Kebbi, Sokoto, Zamfara}; SE={Abia, Anambra, Ebonyi, Enugu, Imo}; SS={Akwa Ibom, Bayelsa, Cross River, Delta, Edo, Rivers} ; SW={Ekiti, Lagos, Ogun, Ondo, Osun, Oyo}. The excluded zone is SW.

the Similar-Ethnicity variable with the index of ethno-linguistic fractionalisation, measured at the state level and the local government level, respectively. Models 1 to 4 use measures of ELF that are calculated at the State Level; while Models 5 to 8 use measures of ELF that are calculated at the local government level. Models IV and VIII are most comparable to the results in Table 6.6 because they include zonal dummies.

No significant results appear when using ELF at the state level, this result remains whether using robust standard errors (Model 1) or clustering them by region (Models 2 to 4). The coefficients only become significant when local government ELF is used. This suggests that a more disaggregated approach is required when looking at the relationship between corruption and ELF. In other words, the ethnic mix of the local government area is more important than the ethnic mix of the state when looking at a firms propensity to bribe, and (potentially) other firm outcomes. Political activities in Nigeria are organised at the National, State and

Table 6.7: Tests For The Equality Of Coefficients Across Firm Level Models With Zonal Dummies

Dependent Variable:	Dummy=1 if firm admitted to bribing; 0 otherwise		
	I	II	χ^2_1 Test Statistic [P-Value]
Similar Ethnicity		-0.147*	
		[0.086]	
Trade	1.132***	1.119***	1.83
	(0.149)	(0.149)	[0.1758]
Employees=50-100	0.148	0.139	1.21
	(0.122)	(0.123)	[0.2707]
Employees=100-250	-0.187	-0.183	0.26
	(0.142)	(0.143)	[0.6132]
Employees=Over 250	-0.390**	-0.378**	1.42
	(0.153)	(0.153)	[0.2328]
Zone==NC	-0.208*	-0.260**	2.69
	(0.117)	(0.121)	[0.1011]
Zone==NE	-0.634***	-0.659***	2.02
	(0.165)	(0.167)	[0.1550]
Zone==NW	-0.561***	-0.568***	0.68
	(0.127)	(0.127)	[0.4088]
Zone==SE	-0.337***	-0.302**	2.71
	(0.116)	(0.118)	[0.1000]
Zone==SS	-0.445***	-0.481***	2.45
	(0.127)	(0.129)	[0.1177]
Constant	-0.369***	-0.267***	2.93
	(0.069)	(0.091)	[0.0870]
Pseudo R-squared	0.072	0.074	
Observations	1267	1267	
Chi-Squared	109.92	112.84	
P-value	0.000	0.000	

$*p < 0.10, **p < 0.05, ***p < 0.01$.Models I-II: Heteroskedasticity Robust Standard Errors In Parenthesis. Similar Ethnicity: Dummy=1 if ethnicity of manager matches ethnicity of LGA Chair, 0 otherwise. Unit of observation is the firm. Similar Ethnicity is calculated using manager data collected by the Author. All other variables are calculated using data from the NBS/EFCC survey. "Zone==..." is a dummy equal to 1 if the firm is located in said geo-political zone, 0 otherwise. Zones are defined as: NC= {Benue, Kogi, Kwara, Nassarawa, Niger, Plateau, FCT (Abuja)}; NE={Adamawa, Bauchi, Borno, Gombe, Taraba, Yobe}; NW={Jigawa, Kaduna, Kano, Katsina, Kebbi, Sokoto, Zamfara}; SE={Abia, Anambra, Ebonyi, Enugu, Imo}; SS={Akwa Ibom, Bayelsa, Cross River, Delta, Edo, Rivers} ; SW={Ekiti, Lagos, Ogun, Ondo, Osun, Oyo}.

Local government level. It appears to be the case that Ethnic networks at the Local Government level are the most important when investigating petty corruption. amongst firms.

Models 5 and 6 model the effect of local government level ELF on the firms propensity to bribe, controlling for ethnic networks. The models are calculated using robust (model 5) and local government area clustered (model 6) standard errors, respectively. Consistent with the macro literature, ethno-linguistic fractionalisation is positively associated with corruption. In model 5, the marginal effect of ELF is 0.2, so a 0.1 increase in the ELF of an area is associated with a 0.02 percentage point increase in the probability of paying a bribe. This result remains significant in model 6 when clustering standard errors by local government area. The coefficient on Similar-Ethnicity is negative in models 5 and 6, however,

Table 6.8: Firm Level Probit Estimations Of Bribery On Matching And ELF, With Zonal Dummies

Dependent Variable:	Dummy=1 if firm admitted to bribing; 0 otherwise							
	1	2	3	4	5	6	7	8
Similar Ethnicity	-0.122	-0.122	-0.486	-0.480*	-0.004	-0.004	-0.283*	-0.373**
	(0.086)	(0.079)	(0.379)	(0.251)	(0.087)	(0.083)	(0.168)	(0.158)
Ethnic Fractionalisation (State Level)	-0.012	-0.012	-0.544	0.378				
	(0.156)	(0.253)	(0.548)	(0.703)				
Similar Ethnicity * (State Level) ELF			0.846	0.869				
			(0.849)	(0.540)				
Ethnic Fractionalisation (LG Level)					0.465***	0.465*	0.040	0.293
					(0.149)	(0.249)	(0.298)	(0.304)
Similar Ethnicity * (LG Level) ELF							0.666*	0.725**
							(0.358)	(0.328)
Zone==NC				-0.552				-0.402**
				(0.465)				(0.177)
Zone==NE				-0.972**				-0.804***
				(0.434)				(0.188)
Zone==NW				-0.535				-0.547***
				(0.369)				(0.178)
Zone==SE				-0.008				-0.067
				(0.325)				(0.197)
Zone==SS				-0.873*				-0.688***
				(0.464)				(0.193)
Constant	-0.481***	-0.481**	-0.209	-0.233	-0.723***	-0.723***	-0.508***	-0.268
	(0.102)	(0.231)	(0.439)	(0.320)	(0.098)	(0.117)	(0.157)	(0.168)
Zones	-	-	-	0.1158	-	-	-	0.0138
Pseudo R-squared	0.002	0.002	0.006	0.042	0.008	0.008	0.011	0.046
Observations	1267	1267	1267	1267	1267	1267	1267	1267
Chi-Squared	2.40	2.36	2.75	11.53	12.58	3.94	6.31	29.49
P-value	0.301	0.307	0.432	0.173	0.002	0.140	0.097	0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$. Models 1 and 5 : Heteroskedasticity Robust Standard Errors In Parenthesis. Models 2, 3 and 4: Standard Errors Clustered By State. Models 6, 7 and 8: Standard Errors Clustered By Local Government Area. Similar Ethnicity: Dummy=1 if ethnicity of manager matches ethnicity of LGA Chair, 0 otherwise. Unit of observation is the firm. Similar Ethnicity is calculated using manager data collected by the Author. All other variables are calculated using data from the NBS/EFCC survey. "Zone==..." is a dummy equal to 1 if the firm is located in said geo-political zone, 0 otherwise. Zones are defined as: NC= {Benue, Kogi, Kwara, Nassarawa, Niger, Plateau, FCT (Abuja)}; NE={Adamawa, Bauchi, Borno, Gombe, Taraba, Yobe}; NW={Jigawa, Kaduna, Kano, Katsina, Kebbi, Sokoto, Zamfara}; SE={Abia, Anambra, Ebonyi, Enugu, Imo}; SS={Akwa Ibom, Bayelsa, Cross River, Delta, Edo, Rivers} ; SW={Ekiti, Lagos, Ogun, Ondo, Osun, Oyo}.The base category for "Zones" is "SW". "Zones" presents the P-value from a Likelihood ratio tests of the equality of the coefficients on the regional effects.

in neither models does the coefficient on Similar-Ethnicity enter significantly.

Models 7 and 8 allow for more flexibility in the analysis by interacting Similar-Ethnicity with LG level ELF. This allows one to see whether the effect of ELF on bribery depends on the ethnic networks in existence in the area. In models 7 and 8 the coeffcient on Similar-Ethnicity is negative and significant; and the coefficient on ELF is positive but no longer significant. However, the coefficient on the interaction term of Similar-Ethnicity and local government ELF is positive and significant. This suggests that a high enough level of fractionalisation will eradicate the advantage of sharing the same ethnicity as the local government politician. The ELF index for Lagos is 0.382. Using this figure along with the results for

Model 7, the impact effect of having a shared ethnicity on the latent propensity to bribe is calculated as: $-0.283 + (0.382 \times 0.666) = -0.029$ which is 90 percent smaller than the 0.283 coefficient on the Similar-Ethnicity variable. This adds more weight to the argument that within a diverse region racial discrimination in the payments of bribes is significantly reduced. Further specifications that allow for ELF to influence bribery unconditional on Similar Ethnicity are included in Table A.89 of Appendix A.4 Section A.4.3. These results seem to tell the same story as above: LG ELF, but not state level ELF, has a positive and significant effect on the propensity to bribe when conditioning on observables; the same applies to religious fracionalisation.

The dataset shows a high correlation, at the state-level and at the local government level, between ELF and the average number of languages spoken within the region. This suggests a possible mechanism through which the causal effect might run. In ethnically mixed areas, people become more used to doing business with other ethnic groups. This might lead people to learn the languages of the other ethnic groups; this polylingualism might help in building rapport between people of different ethnic groups which can reduce ethnic discrimination in bribe payments. The simple correlation between ELF and the average number of languages spoken by managers is 0.61 (0.35) and is statistically significant at the 1% level.

Table 6.9 investigates the nature of bribery in the South-Western zone of the country by running probit estimations on the firms in this zone alone. Regional dummies are included in models 4 to 6 with Lagos State being the omitted region. All regional dummies have negative coefficients that are statsitically significant, which suggests that on average and ceteris paribus, firms in Lagos are more likely to pay a bribe than firms in other states in the South-West region. Out of the states in the South-West region, Lagos is the most ethnically diverse and also has the highest indices of ethnic fractionalisation and polarisation. This coheres with the argument that shared ethnicities only matter for bribery in more ethnically homogeneous settings. Lagos also has the least percentage of cases of shared ethnicities (56%); the state with the next highest percentage of matches is Oyo State (80%). This is probably why the Similar-Ethnicity variable loses significance when controlling for the states. These results are consistent with those in Tables 6.6 and 6.8: the incidence of bribery is higher in Lagos; and in ethnically mixed environments, sharing the same ethnicity as the local politician will not have as much of an effect on the propensity to bribe compared with areas that are less ethnically fractionalised.

To further investigate the proposition that bribery only takes on a racial dimension in racially uniform areas, Table 6.10 presents result from models focusing solely on Lagos state, with local government dummies in models 4 to 6. The

Similar-Ethnicity variable does not enter significantly in any of the models and 13 of the 15 local government dummies are not statistically different from zero (the base local government is Ikeja). Taken in full these results suggest that, conditional on covariates, whilst there is variation in bribe payments and a significant effect of ethnic discrimination on bribe payments across Zones and across States within the South-Western zone, there is little evidence for these effects within Lagos itself. Sharing the same ethnicity as the local government politician does not seem to have any benefit within Lagos State. These results are corroborated in Table 6.11 which controls for zonal heterogeneity in the Similar-Ethnicity variable. Having the same ethnicity as the local politician reduces the chances of paying a bribe, on average and ceteris paribus, in the South-Western Zone.

Table 6.9: South-West Zone: Firm Level Probit Estimations Of Bribery On Matching, With Regional Dummies

Dependent Variable:	Dummy=1 if firm admitted to bribing; 0 otherwise					
	1-SW	2-SW	3-SW	4-SW	5-SW	6-SW
Similar Ethnicity	-0.508***	-0.452***	-0.424***	-0.003	0.013	0.036
	(0.131)	(0.134)	(0.137)	(0.146)	(0.148)	(0.151)
Trade		0.768***	0.839***		0.427*	0.501**
		(0.205)	(0.211)		(0.228)	(0.234)
Employees=50-100			0.030			-0.111
			(0.191)			(0.203)
Employees=100-250			-0.256			-0.415*
			(0.213)			(0.236)
Employees=Over 250			-0.402*			-0.416*
			(0.223)			(0.248)
State Dummies	NO	NO	NO	YES^{--}	YES^{--}	YES^{--}
Constant	0.065	-0.057	-0.027	0.210*	0.133	0.206
	(0.108)	(0.114)	(0.123)	(0.112)	(0.120)	(0.131)
Pseudo R-squared	0.026	0.051	0.058	0.233	0.241	0.249
Observations	439	439	439	439	439	439
Chi-Squared	14.96	27.85	32.10	94.71	110.04	119.31
P-value	0.000	0.000	0.000	0.000	0.000	0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$. $^{++}(^{--})$: Plusses (minuses) indicate that the coefficients on all state dummies are positive (negative) and significant at, at least, the 5% level. Heteroskedastic Robust Standard Errors In Parenthesis. Similar Ethnicity: Dummy=1 if ethnicity of manager matches ethnicity of LGA Chair, 0 otherwise. Unit of observation is the firm. Similar Ethnicity is calculated using manager data collected by the Author. All other variables are calculated using data from the NBS/EFCC survey. These estimations are run on firms that are located in the South-West Zone only.

6.5.2 Robustness Checks

Models I to IV Table 6.8 also shows results from models that use State-Level ELF as an explanatory variable instead of Local Government level ELF. Using ELF defined at the state level fails to bring up any statistically significant results, suggesting that it is the local area which is important in determining the nature of ethnic discrimination in bribe payments, rather than the wider economy. Ag-

Table 6.10: Lagos: Firm Level Probit Estimations Of Bribery On Matching, With Local Government Area Dummies

Dependent Variable:	Dummy=1 if firm admitted to bribing; 0 otherwise					
	1-Lagos	2-Lagos	3-Lagos	4-Lagos	5-Lagos	6-Lagos
Similar Ethnicity	-0.026	-0.015	0.004	-0.058	-0.045	-0.013
	(0.155)	(0.155)	(0.158)	(0.166)	(0.167)	(0.170)
Trade		0.264	0.321		0.203	0.278
		(0.212)	(0.218)		(0.220)	(0.224)
Employees=50-100			-0.074			-0.121
			(0.222)			(0.233)
Employees=100-250			-0.313			-0.360
			(0.254)			(0.261)
Employees=Over 250			-0.282			-0.392
			(0.281)			(0.295)
LG==agege				0.119	0.108	-0.001
				(0.460)	(0.462)	(0.472)
LG==ajeromi_ifelodun				0.238	0.235	0.134
				(0.604)	(0.597)	(0.599)
LG==alimosho				-0.028	-0.012	-0.114
				(0.337)	(0.336)	(0.344)
LG==apapa				0.291	0.282	0.289
				(0.371)	(0.368)	(0.377)
LG==eti_osa				0.671**	0.665**	0.620**
				(0.304)	(0.306)	(0.310)
LG==ifako_ijaye				-0.171	-0.190	-0.241
				(0.351)	(0.354)	(0.357)
LG==ikorodu				0.086	0.103	0.062
				(0.399)	(0.398)	(0.397)
LG==kosofe				-0.597	-0.574	-0.608
				(0.402)	(0.405)	(0.404)
LG==lagos_island				0.417	0.421	0.359
				(0.386)	(0.390)	(0.402)
LG==lagos_mainland				0.153	0.160	0.135
				(0.386)	(0.387)	(0.398)
LG==mushin				0.546	0.556	0.538
				(0.540)	(0.537)	(0.515)
LG==ojo				0.802	0.840	0.914
				(0.669)	(0.670)	(0.623)
LG==oshodi_isolo				-0.267	-0.240	-0.198
				(0.360)	(0.360)	(0.359)
LG==shomolu				0.429	0.430	0.401
				(0.567)	(0.563)	(0.541)
LG==surulere				0.664**	0.657**	0.690**
				(0.304)	(0.304)	(0.310)
Constant	0.223*	0.174	0.224*	0.060	0.017	0.101
	(0.116)	(0.123)	(0.135)	(0.209)	(0.214)	(0.232)
Pseudo R-squared	0.000	0.004	0.010	0.058	0.061	0.069
Observations	273	273	273	270	270	270
Chi-Squared	0.03	1.60	3.71	20.39	20.75	25.04
P-value	0.864	0.449	0.593	0.203	0.238	0.200

$*p < 0.10, **p < 0.05, ***p < 0.01$. Heteroskedastic Robust Standard Errors In Parenthesis. Similar Ethnicity: Dummy=1 if ethnicity of manager matches ethnicity of LGA Chair, 0 otherwise. Unit of observation is the firm. Similar Ethnicity is calculated using manager data collected by the Author. All other variables are calculated using data from the NBS/EFCC survey. These estimations are run on firms that are located in Lagos State only.

gregating things up to the state level seems to wash away all of the significant effects.

Models I and II present the same model using heteroskedasticity robust standard errors; and standard errors clustered by region, respectively. Calculating the standard errors of the coefficients differently does not seem to alter the results significantly. None of the coefficients in either model are statistically significant.

The same results seem to occur when using either NBS or CLEEN dataset to calculate the index of ethno-linguistic fractionalisation. Figure 16 shows the ethnic composition of the different regions in Nigeria (Yoruba; Ibo; Hausa; and

Table 6.11: Firm Level Probit Estimations Of Bribery On Matching, Controlling For Zonal Heterogeneity

Dependent Variable:	Dummy=1 if firm admitted to bribing; 0 otherwise			
	1	2	3	4
Similar Ethnicity	-0.194**	0.025	-0.508***	0.177
	(0.084)	(0.109)	(0.131)	(0.207)
Zone==SW		0.065		0.065
		(0.108)		(0.108)
Similar Ethnicity*Zone==NC			0.685***	
			(0.245)	
Similar Ethnicity*Zone==NE			-0.143	-0.828**
			(0.341)	(0.376)
Similar Ethnicity*Zone==NW			0.656**	-0.028
			(0.263)	(0.308)
Similar Ethnicity*Zone==SE			0.622*	-0.063
			(0.373)	(0.406)
Similar Ethnicity*Zone==SS			0.555**	-0.130
			(0.253)	(0.300)
Similar Ethnicity*Zone==SW		-0.533***		-0.685***
		(0.171)		(0.245)
Constant	-0.148*	-	0.065	-
	(0.084)	-	(0.108)	-
Zonal Dummies	YES^{--}	YES^{--}	YES^{--}	YES^{--}
Zones	0.0428	0.0000	0.6267	0.0000
Interactions	-	-	0.2375	0.0217
Pseudo R-squared	0.031		0.042	
Observations	1267	1267	1267	1267
Chi-Squared	46.09	269.85	60.51	270.49
P-value	0.000	0.000	0.000	0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$. $^{++}(^{--})$: Plusses (minuses) indicate that the coefficients on all zonal dummies are positive (negative) and significant at, at least, the 5% level. Models 1-4: Heteroskedastic Robust Standard Errors In Parenthesis. Similar Ethnicity: Dummy=1 if ethnicity of manager matches ethnicity of LGA Chair, 0 otherwise. Unit of observation is the firm. Similar Ethnicity is calculated using manager data collected by the Author. All other variables are calculated using data from the NBS/EFCC survey. "Zone==..." is a dummy equal to 1 if the firm is located in said geo-political zone, 0 otherwise. Zones are defined as: NC={Benue, Kogi, Kwara, Nassarawa, Niger, Plateau, FCT (Abuja)}; NE={Adamawa, Bauchi, Borno, Gombe, Taraba, Yobe}; NW={Jigawa, Kaduna, Kano, Katsina, Kebbi, Sokoto, Zamfara}; SE={Abia, Anambra, Ebonyi, Enugu, Imo}; SS={Akwa Ibom, Bayelsa, Cross River, Delta, Edo, Rivers} ; SW={Ekiti, Lagos, Ogun, Ondo, Osun, Oyo}.The base category for "Zones" is "SW". "Zones"; and "Interactions" present the P-value from Likelihood ratio tests of the equality of the coefficients on the zonal effects and interaction terms, respectively.

Other) using the two datasets, respectively. The two datasets seems to generate the same results. Both datasets seem to tell a similar story: the Northern region is mainly populated by the Hausa tribe; the South-Eastern region by the Ibo and Other tribes; and the South-Western region by the Yoruba tribe.

Table 6.12 includes a dummy ("Similar-Language") indicating whether or not the firm manager spoke the same language as the most common ethnic group. The data on languages was collected by the author; the dummy variable is based on the survey question that asked the managers what questions they spoke. Using this data has the advantage that languages spoken is not subject to the error that occurs when language is inferred from ethnicity. This study is able to identify which languages are spoken by the manager, as reported by the manager, and use this to see whether there is any advantage of speaking the same language as the politician when conducting operations where a bribe might be extracted from the manager.

The first model is the same as Model 4 of Table 6.6. The second model substitutes Similar-Language for Similar-Ethnicity. The third model includes both dummies in the estimation. Using Similar-Language instead of Similar-Ethnicity (Model 2) seems to tell the same story, managers who speak the same language as the majority group are, on average and ceteris paribus, less likely to pay a bribe. When both dummies are included in the estimation, the coefficients retain their sign but lose statistical significance, giving more weight to the argument that the Similar-Ethnicity variable is driven by the level of ethnic and linguistic diversity within the regions.

Table 6.12: Robustness Checks Using Languages Spoken By The Firm Managers

Dependent Variable:	Dummy=1 if firm admitted to bribing; 0 otherwise		
	1	2	3
Similar Ethnicity	-0.194**		-0.128
	(0.084)		(0.097)
Similar-Language		-0.232**	-0.154
		(0.099)	(0.115)
Zonal Dummies	YES^{--}	YES^{--}	YES^{--}
Constant	-0.148*	-0.087	-0.065
	(0.084)	(0.103)	(0.104)
Pseudo R-squared	0.031	0.031	0.033
Observations	1267	1267	1267
Chi-Squared	46.09	46.81	48.23
P-value	0.000	0.000	0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$. $^{++}(^{--})$: Plusses (minuses) indicate that the coefficients on all zonal dummies are positive (negative) and significant at, at least, the 5% level. Heteroskedastic Robust Standard Errors In Parenthesis. Similar Ethnicity: Dummy=1 if ethnicity of manager matches ethnicity of LGA Chair, 0 otherwise. Similar-Language: Dummy=1 if the firm manager speaks the same language spoken by the most common ethnic group, 0 otherwise. Unit of observation is the firm. Similar Ethnicity and Similar-Language are calculated using manager data collected by the Author. All other variables are calculated using data from the NBS/EFCC survey. "Zone==..." is a dummy equal to 1 if the firm is located in said geo-political zone, 0 otherwise. Zones are defined as: NC= {Benue, Kogi, Kwara, Nassarawa, Niger, Plateau, FCT (Abuja)}; NE={Adamawa, Bauchi, Borno, Gombe, Taraba, Yobe}; NW={Jigawa, Kaduna, Kano, Katsina, Kebbi, Sokoto, Zamfara}; SE={Abia, Anambra, Ebonyi, Enugu, Imo}; SS={Akwa Ibom, Bayelsa, Cross River, Delta, Edo, Rivers} ; SW={Ekiti, Lagos, Ogun, Ondo, Osun, Oyo}.The base category for "Zones" is "SW".

6.6 Conclusion

This study set out to investigate the links between ethnicity and bribery amongst firm managers in Nigeria. It is the first analysis of its kind that identifies the ethnic groups of firm managers; households; and local government politicians. Using this information this study constructs indices of ethnolinguistic fractionalisation and measures of ethnic networks.

The main questions that were investigated were: whether belonging to the ethnic network of a local politician had a positive or negative effect on the probability of paying a bribe; whether or not this impact varied based on the level of ethno-linguistic fractionalisation within the area; and whether these factors were related to victimisation and the reporting of crime.

The problem of acquiring individual-level data on ethnicities was overcome by contacting the firm-managers directly and asking them for their ethnic group, state of origin, religion, and the languages that they had knowledge of. Use of this data allowed for the construction of variables measuring ethnic networks, linguistic networks, ethno-linguistic fractionalisation; linguistic fractionalisation; and religious fractionalisation. Measures of polarisation were also used in the analysis (See Appendix A.4 Section A.4.3). The study used information concerning over 90 ethnic groups to construct weighted and unweighted measures of ethno-linguistic fractionalisation. These encorporate the phylogenetic distances between the ethnic groups in the sample so that the difference between two ethnic groups will depend on their historical origin. So, for example, Yoruba and Igbo are treated as more closely related than Yoruba and Hindu.

Results showed evidence that belonging to the same ethnic network as the local politician significantly reduced the probability of paying a bribe. This result was robust to a series of model specifications and occurred for the entire sample of Nigeria; and a sub-sample of south-western states. This result disappeared, however, within Lagos State. Further analysis suggested that the effect of ethnic networks is only significant in areas that have low levels of ethnic diversity. In cosmopolitan areas (like Lagos) the effect of ethnic networks on bribery ceases to exist. This result corresponds with previous work on discrimination in areas with ethnic minorities.

Using information concerning the languages spoken by the managers and the most common ethnic group in the local government area, this study tested whether speaking the same language as the local politician had any effect on the propensity to bribe. Results showed that if a manager spoke the same language as a politician then they were less likely to pay a bribe, ceteris paribus. However, this study was unable to disentangle this effect from the effect of sharing the same ethnicity as the politician. 85% of all managers who were non-speakers of the local language did not share the same ethnicity as the politicians; while 77% of speakers shared the same ethnicity as the local politician. The two variables are positively corelated and this relationship is statistically significant at the 10% level. Future research can try to disentangle the effect of speaking the same language from the effect of sharing the same ethnic group.

The effect of ethno-linguistic fractionalisation on corruption was only found to be statistically significant at the local government level. State-level estimations returned insignificant coefficients. The effect of sharing an ethnicity on the propensity to bribe was stronger in the local government level estimations compared with the state-level estimations.

The study was limited by the use of a cross-section, nevertheless, arguments were provided for the exogeneity of a shared ethnicity on the propensity to bribe. The ethnic group that the local politician comes from, in a given area, is stable and the majority of firms are small-sized family run firms, so the owner of the firm and the top manager are usually the same person. Therefore, the ethnicity of the manager is also stable. The median firm age was 14 years; which suggests that firms in this sample are not re-locating based on frequent changes in local politicians. Future research could investigate whether managers are likely to migrate to areas where the local politician shares their ethnicity.

7 Discussion

The aim of this dissertation was to discuss the nature and workings of firm level corruption in Nigeria. The analysis focused on the payment of bribes from firm managers to local public officials. Central to this empirical study of bribery was the topic of measurement error and how one knows that the reports of firm managers accurately capture their experiences. This study dwelt on this issue by looking at different ways of measuring firm-level corruption; and the determinants of underreporting of corruption by managers. The determinants of firm-level bribery were investigated with a focus on firm-specific and public official characteristics. The first and fourth main chapters focused on the firm and managerial specific factors influencing the payment of bribery, respectively. The second and third main chapters looked at the self-reporting and underreporting of corruption and business malpractices.

Contributions to the body of knowledge include: the application of existing models of corruption to a new setting: West Africa; an increase in the number of variables that capture public officials meeting with firm managers; a relaxation of the assumption that public officials extort bribes from firms, this was done by examining the extent to which firms offer bribes in return for a favour; a comparison of the randomised response and indirect questioning methods in measuring corruption; an analysis of the causes of underrepoting by firm managers in the presence of statistical and assured anonymity; an analysis of the impact of ethnic networks in firm level dealings with the government; and the introduction of individual, firm, local government, and state level data in the study of ethnicity and corruption.

The investigation began with a discussion of the nature of corruption at the firm level in Nigeria. While still categorised as more corrupt than the median country, according to Transparency International's Corruption Perceptions Index, the level of corruption in the country seems to be decreasing. Much of this change has been attributed to the institutions of anti-corruption laws and bodies within the country. Recent measures of corruption rank the level of corruption in the country as being similar to that of Indonesia and Kenya.

The first main chapter looked at the determinants of the incidence and level of bribe payments amongst manufacturing firms in Nigeria. This study addressed the issue of the underreporting of corruption by using data from surveys that asked managers about their bribery directly and indirectly. Information from two datasets was used independently of each other. Nevertheless, the results were consistent across both datasets. The determinants of the incidence of bribery were found to be different to the factors that influenced the size of bribe paid. The former was determined by the meetings with public officials while the latter was determined by firm specific characteristics such as profits and capital mobility.

Also, meetings with officials from the police was found to be the strongest determinant of bribery compared to other organisations. This result is significant for the literature concerning: the detection of corruption; and the costs of corruption. Knowing where corruption is most likely to take place can assist in measuring it. This knowledge can also be used to change the incentives of firm managers by running targetted anti-corruption efforts. Future research can use a larger sample to examine the propensity to pay a bribe to an agent conditional on meeting that type if agent.

The second main chapter investigated the validity of using indirect questioning methods to measure corruption amongst firms. These methods were compared to the randomised response method which attaches responses from direct questions to a probability distribution in order to assure anonymity while also allowing for an estimate of the mean rate of corruption. The comparison of the point estimates from the indirect and randomised response questioning methods comprised a gap in the litarature that compares the estimates from the indirect; direct; and randomised response methods. The results of this study (Chapter 4) show that indirect questioning methods are preferred to randomised question methods; moreover, despite the assurance of anonymity, evidence is found for a significant proportion of managers choosing to misreport (underreport) their true status. The third main chapter seeks to answer why this is the case. Traditional research suggests that individuals misreport their sensitive behaviour to adhere to social norms. However, this chapter addresses why anonymity does not eliminate this bias. A host of reasons are investigated, including : education; understanding of the question; profits; guilt; and interviewer & supervisor effects. None of these variables were significant in explaining why managers chose to underreport their status. Nevertheless, evidence was found for trust and the probability/fear of detection being closely associated with underreporting (reticence) amongst managers. Trust was treated as a latent variable which was indicated by the nature of business contracts that the managers engaged in; fear of detection was measured by the perceived number of close competitors in the same business as the firm. These results suggest that more accurate sensitive data can be achieved from larger and less differentiated populations. These results are significant for the literature on measurement error in corruption; and survey research methods. The information concerning the determinants of reticence can assist in future survey research on sensitive questions. Future research can examine these results experimentally by monitoring results from a random event, attaching the event to a sensitive questionnaire, and monitoring the misreporting behaviour of the managers. Furthermore, a comparison of the variances derived from the randomised response and indirect questioning methods can also fill a gap in the literature.

The final main chapter focuses on the links between ethnic networks and corruption. This is done by looking at the ethnicities of firm managers of 1,267 firms; and the local politicians in 351 local government areas in Nigeria. The study investigated the effect on bribery of sharing the same ethnicity as the local politicians. It was argued that firm managers who belong to the same ethnic group as the local politician could be less likely to pay a bribe due to ethnic discrimination of non co-ethnics; and/or punishment mechanisms existing within one's kin group. Alternatively they might be more likely to bribe due to revenue sharing among one's kin group. This study also constructed the first set of micro-level measures of ethno-linguistic fractionalisation (ELF) for Nigeria and investigated the impact of ELF on firm level bribery. Results showed that firms whose ethnicity matched that of the local politicians were less likely to pay a bribe. The findings also confirmed the results of the macro literature: ELF, measured at the local government level, is positively and significantly associated with a higher rate of bribery. This result only shows up at the local government level and not at the state level, suggesting that local government level activity is more important when looking at the effect of ethnic networks on firm level behaviour relating to corruption. Another result came from running the analysis with a Similar-Ethnicity-ELF interaction term. Results from these models showed that with a high enough level of ELF, the negative impact of having a co-ethnic politician on bribery disappears. This suggests that ethnic discrimination in the payment of bribes is an issue for areas that lack ethnic diversity. In places, such as Lagos, with a relatively high level of ELF the effect of ethnic networks on the propensity to bribe goes away. These results were used alongside results from the economics of ethnic discrimination to suggest that such discrimination is more likely to occur in ethnically uniform societies. This piece of work can be of importance to the literature on discrimination, in general, and ethnic discrimination in particular. An understanding of the impact of fractionalisation on discrimination can assist in detecting where discrimination is likely to have its greatest effect. Future research can examine the impact of ethnic networks on other firm outcomes such as profitability and market capitalisation. In addition, work can be done to see whether the effect of discrimination is affected by the level of fractionalisation for different types of discrimination.

A Appendices

A.1 Appendices For Chapter 1

A.1.1 Bivariate Probit Model With Sample Selection

The following tables show the results of the Bivariate Probit Analysis of the bribery of public officials.

Table A.1: Information On Actions Performed By Companies Which Required Them To Meet With Government Officials

Dataset	ES	NBS
Actions	• Requested a mainline telephone connection in the last 2 years • Requested an electrical connection in the last 2 years • Requested a water connection in the last 2 years • Requested a construction-related permit in the last 2 years • Requested an import license in the last 2 years • Requested an operating license in the last 2 years • Visited by, inspected by, or required to meet with tax officials in the past year	• Cleared Goods Through Customs • Obtained Road Worthy Certificates • Procurement Of Goods And Services From Government • Obtaining Business Licenses And Permits • Procurement Of Goods And Services From Private Companies • Getting Clearance For Environmental Or Sanitary Regulations • Residence And Work Permits • Vehicle Registrations • Police Investigations • Traffic Offences • Contact With The Court

Table A.2: Summary Statistics For Meeting With Government Officials And Bribing Government Officials

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dataset	ES						NBS					
	Manufacturing			Full Sample			Manufacturing			Full Sample		
	All Firms	Met With An Official	Bribed An Official	All Firms	Met With An Official	Bribed An Official	All Firms	Met With An Official	Bribed An Official	All Firms	Met With An Official	Bribed An Official
Met With An Official	0.92	1	1	0.92	1	1	0.37	1	1	0.31	1	1
(Mean) Average Number Of Types Of Officials Met With	1.68 (1.22)	1.83 (1.17)	2.26 (1.30)	1.88(1.34)	2.04 (1.27)	2.48 (1.33)	1.66 (2.72)	4.43 (2.74)	4.2 (2.50)	1.09 (2.07)	3.49 (2.32)	3.49 (2.33)
Bribery Episode	0.35	0.38	1	0.40	0.43	1	0.14	0.36	1	0.10	0.33	1
(Mean) Number Of Types Of Bribery Episodes	0.56 (0.95)	0.61 (0.97)	1.59 (0.95)	0.69 (1.07)	0.75 (1.09)	1.72 (1.04)	0.44 (1.39)	1.16 (2.07)	3.2 (2.31)	0.28	0.90 (1.76)	2.67 (2.12)
Observations	2001	1839	710	3668	3377	1463	331	124	45	2089	651	218

Table A.3: Bivariate Probit Model With Sample Selection - Estimations On Manufacturing (ES)

	telecoms	electric	water	construction	import	tax
Outcome Equation						
foreign	.	.	.	.	.	-16.988
	.	.	.	.	.	(0.000)
security	-0.126	0.213**	0.243	0.103	0.398**	0.228***
	(0.153)	(0.105)	(0.223)	(0.225)	(0.158)	(0.071)
age	0.001	0.001	0.008	-0.002	0.007	-0.002
	(0.009)	(0.006)	(0.013)	(0.013)	(0.009)	(0.004)
trade	0.248	0.122	0.430	0.573	0.251	0.150
	(0.310)	(0.302)	(0.870)	(0.598)	(0.344)	(0.206)
Constant	-1.610***	-0.725***	-0.696	1.322*	-0.785	-0.887***
	(0.212)	(0.175)	(0.486)	(0.770)	(0.599)	(0.082)
Selection Equation						
Industry Dummies	YES	YES	YES	YES	YES	YES
Regional Dummies	YES	YES	YES	YES	YES	YES
Constant	-1.366***	-0.494**	-0.650***	-1.214***	-0.033	0.230
	(0.244)	(0.199)	(0.240)	(0.233)	(0.205)	(0.179)
Observations	2001	2001	2001	2001	2001	2001
Censored Observations	1759	1372	1759	1837	1599	332
Log-Likelihood	-734.5	-1448.9	-753.4	-598.3	-1094.2	-1686.9
Chi-Squared	1.3	4.3	1.9	1.0	7.9	.
P Value	0.732	0.228	0.590	0.809	0.048	.
Rho	0.766	0.441	-0.052	-0.833	0.184	0.943
Wald (Rho=0)	6.71	5.57	0.04	0.76	0.13	8.02
Prob>chi2	0.0096	0.0183	0.8512	0.3827	0.7231	0.0046

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dependent Variable in selection equation is Dummy=1 if firm requested public service (Table A.1), 0 otherwise. Dependent Variable in outcome equation is Dummy=1 if a bribe was requested/expected from the firm, 0 otherwise. (Huber-White) Heteroscedastic Robust standard errors in parenthesis. The base group for state-dummies is Abia state. The base group for sector dummies is "Other - Manufacturing".

Table A.4: Bivariate Probit Model With Sample Selection - Estimations On All Firms (ES)

	telecoms	electric	water	construction	import	tax
Outcome Equation						
foreign	0.236	-0.387	-4.534***	-5.278***	-0.329	-0.782
	(0.635)	(0.740)	(0.255)	(1.766)	(0.723)	(0.503)
security	0.191*	0.293***	-0.006	0.223	0.215**	0.146***
	(0.100)	(0.074)	(0.129)	(0.183)	(0.106)	(0.051)
age	-0.001	0.008	-0.003	0.001	0.001	0.001
	(0.007)	(0.005)	(0.009)	(0.012)	(0.006)	(0.003)
trade	-0.036	-0.114	0.359	0.813	0.003	0.074
	(0.285)	(0.315)	(0.907)	(0.669)	(0.321)	(0.194)
Constant	-1.498***	-0.465***	-0.634*	0.312	-0.519**	-0.832***
	(0.168)	(0.179)	(0.362)	(2.246)	(0.251)	(0.057)
Selection Equation						
Industry Dummies	YES	YES	YES	YES	YES	YES
Regional Dummies	YES	YES	YES	YES	YES	YES
Constant	-1.070***	-0.271*	-1.031***	-1.404***	-0.057	0.082
	(0.163)	(0.150)	(0.179)	(0.228)	(0.158)	(0.135)
Observations	3668	3668	3668	3668	3668	3668
Censored Observations	3031	2226	3128	3307	2794	706
Log-Likelihood	-1879.6	-3184.0	-1707.7	-1346.5	-2403.4	-3393.5
Chi-Squared	3.9	19.1	324.2	11.7	4.3	10.6
P Value	0.419	0.001	0.000	0.019	0.361	0.031
Rho	0.658	0.106	0.181	-0.286	0.198	0.994
Wald (Rho=0)	10.13	0.28	0.62	0.04	0.93	39.47
Prob > Chi2	0.0015	0.5993	0.4295	0.8385	0.3355	0

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dependent Variable in selection equation is Dummy=1 if firm requested public service, 0 otherwise (Table A.1). Dependent Variable in outcome equation is Dummy=1 if a bribe was requested/expected from the firm, 0 otherwise. (Huber-White) Heteroscedastic Robust standard errors in parenthesis. The base group for state-dummies is Abia state. The base group for sector dummies is “Rest Of Universe - Other”.

Table A.5: Bivariate Probit Model With Sample Selection - Estimations On Manufacturing Companies (NBS)

	customs	road	goods_gov	lic	goods_priv	enviro	res_work	vehicle	police	traffic	court
Outcome Equation											
foreign	0.043	0.244	0.643*	-0.306	0.506	-0.375	-0.201	0.276	-0.156	-0.080	-6.003
	(0.040)	(0.307)	(0.382)	(0.261)	(0.566)	(0.394)	(0.383)	(0.296)	(21.608)	(0.161)	(0.000)
security	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.
age	0.000	0.007	0.012	0.005	0.008	0.009	-0.012	0.007	0.003	0.003	-0.010
	(0.003)	(0.008)	(0.014)	(0.009)	(0.028)	(0.015)	(0.009)	(0.008)	(0.173)	(0.004)	(0.015)
trade	.	-0.211	-0.960**	-0.247	-6.043***	0.068	-0.394	-0.351	0.039	-0.381***	-0.970**
	.	(0.309)	(0.434)	(0.292)	(1.300)	(0.368)	(0.330)	(0.305)	(18.601)	(0.108)	(0.478)
Constant	0.870***	-1.782***	-1.810***	-1.547***	-2.140***	0.072	-1.347***	-1.596***	1.302	-1.153***	-1.115***
	(0.104)	(0.224)	(0.308)	(0.193)	(0.753)	.	(0.185)	(0.175)	(7.254)	(0.065)	(0.426)
Selection Equation											
Regional Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Industry Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Constant	-0.015	-1.822***	-5.789	-1.673***	-5.626***	-0.008	-1.812***	-1.814***	-0.185	-1.448***	-2.286***
	(0.349)	(0.276)	.	(0.250)	(0.109)	(0.505)	(0.327)	(0.287)	(3.467)	(0.076)	(0.372)
Observations	331	331	331	331	331	331	331	331	331	331	331
Censored Observations	274	256	305	272	302	271	281	256	293	273	309
Log-Likelihood	-154.6	-188.1	-87.0	-161.0	-90.2	-168.0	-137.8	-192.0	-118.8	-173.3	-72.3
Chi-Squared	1.8	1.5	.	4.0	.	1.5	8.0	1.5	.	16.2	.
P Value	0.407	0.679	.	0.261	.	0.685	0.045	0.671	.	0.001	.
Rho	-1.000	1.000	1.000	1.000	1.000	-0.544	1.000	1.000	-1.000	1.000	1.000
Wald (Rho=0)	81.41	56.26	9.03	18.46	7.85	1.9	4.43	0.2	0	20.06	108.63
Prob > chi2	0	0	0.0027	0	0.0051	0.1679	0.0353	0.6553	0.9663	0	0

$*p < 0.10, **p < 0.05, ***p < 0.01$

Dependent Variable in selection equation is Dummy=1 if firm met with public official, 0 otherwise (Table A.1). Dependent Variable in outcome equation is Dummy=1 if a bribe was paid to the public official, 0 otherwise. The base group for state-dummies is Abia state.

Table A.6: Bivariate Probit Model With Sample Selection - Estimations On All Firms (NBS)

	customs	road	goods_gov	lic	goods_priv	enviro	res_work	vehicle	police	traffic	court
Outcome Question											
foreign	-0.199	-0.174	0.067	-0.192	0.063	-0.474*	-0.153***	-0.099	-0.366**	-0.195***	-0.232***
	(0.194)	(0.254)	(0.185)	(0.441)	(0.259)	(0.271)	(0.024)	(0.266)	(0.148)	(0.017)	(0.003)
security	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.
age	-0.003	-0.001	-0.004	-0.007	-0.006	-0.000	-0.002**	0.004	-0.000	-0.002***	-0.001***
	(0.005)	(0.005)	(0.005)	(0.007)	(0.005)	(0.006)	(0.001)	(0.005)	(0.003)	(0.000)	(0.000)
trade	.	0.150	0.160	0.243	-0.022	0.161	-0.212***	0.255	-0.032	-0.295***	-0.051***
	.	(0.174)	(0.162)	(0.298)	(0.201)	(0.273)	(0.015)	(0.210)	(0.107)	(0.018)	(0.003)
Constant	-1.530***	-1.500***	-2.102***	-1.457	-2.174***	-0.407	1.478***	-1.702	-1.654***	1.636***	1.489***
	(0.380)	(0.431)	(0.121)	(1.443)	(0.016)	(2.389)	(0.014)	(1.150)	(0.078)	(0.014)	(0.002)
Selection Equation											
Industry Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Regional Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Constant	-1.953***	-2.053***	-1.374***	-1.283***	-1.591***	-1.006**	-1.359***	-1.378***	-1.336***	-1.588***	-1.094***
	(0.286)	(0.276)	(0.232)	(0.416)	(0.248)	(0.393)	(0.239)	(0.224)	(0.183)	(0.115)	(0.003)
Observations	2089	2089	2089	2089	2089	2089	2089	2089	2089	2089	2089
Censored Observations	1935	1821	1950	1813	1942	1852	1917	1756	1877	1880	1964
Log-Likelihood	-554.1	-870.4	-540.8	-895.2	-554.5	-799.9	-612.6	-1024.8	-768.3	-743.1	-468.8
Chi-Squared	1.5	1.0	2.0	1.2	2.2	3.1	459.6	2.1	7.2	904.5	56911.5
P Value	0.467	0.804	0.581	0.746	0.532	0.370	0.000	0.544	0.065	0.000	0.000
Rho	0.795	0.569	1.000	0.617	1.000	-0.111	-1.000	0.558	1.000	-1.000	-1.000
Wald (Rho=0)	3.36	1.62	1.12	0.13	2.94	0.01	36.8	0.14	1.08	24.63	7.38
Prob > chi2	0.0669	0.2036	0.2893	0.7174	0.0866	0.9398	0	0.7053	0.2983	0	0.0066

$*p < 0.10, **p < 0.05, ***p < 0.01$

Dependent Variable in selection equation is Dummy=1 if firm met with public official, 0 otherwise (Table A.1). Dependent Variable in outcome equation is Dummy=1 if a bribe was paid to the official, 0 otherwise. The base group for state-dummies is Abia state. The base group for sector dummies is “Other community, social and personal services”.

Table A.7: Dummies In Secondary Bivariate Probit Model With Sample Selection

ES Dummies	Description	NBS Dummies	Description
ic_telecom	Dummy=1 if telecommunications reported as minor, moderate,major, or very severe obstacle to operations	obs_tax_regs	obstacles for doing good business - tax regulations
ic_elec	Dummy=1 if electricity reported as minor, moderate,major, or very severe obstacle to operations	obs_infl	obstacles for doing good business - inflation
ic_trans	Dummy=1 if transportation reported as minor, moderate,major, or very severe obstacle to operations	obs_pol	obstacles for doing good business - political instability
ic_land	Dummy=1 if access to land for expansion/relocation reported as minor, moderate,major, or very severe obstacle to operations	obs_change_law	obstacles for doing good business - changes in laws and regulations
ic_taxrates	Dummy=1 if tax rates reported as minor, moderate,major, or very severe obstacle to operations	obs_crime	obstacles for doing good business - crime and insecurity
ic_taxadmin	Dummy=1 if tax administration reported as minor, moderate,major, or very severe obstacle to operations	obs_corr	obstacles for doing good business - corruption
ic_customs	Dummy=1 if customs and trade regulations reported as minor, moderate,major, or very severe obstacle to operations	obs_bus_reg	obstacles for doing good business - complicated business registration
ic_laborregs	Dummy=1 if labour regulations reported as minor, moderate,major, or very severe obstacle to operations	obs_unclear_laws	obstacles for doing good business - unclear laws
ic_educlabour	Dummy=1 if inadequately educated workforce reported as minor, moderate,major, or very severe obstacle to operations		
ic_buslic	Dummy=1 if business licensing and permits reported as minor, moderate,major, or very severe obstacle to operations		
ic_accfin	Dummy=1 if access to finance (e.g. collateral) reported as minor, moderate,major, or very severe obstacle to operations		
ic_costfin	Dummy=1 if cost of finance reported as minor, moderate,major, or very severe obstacle to operations		
ic_polenviro	Dummy=1 if political environment reported as minor, moderate,major, or very severe obstacle to operations		
ic_macroenviro	Dummy=1 if macroeconomic environment reported as minor, moderate,major, or very severe obstacle to operations		
ic_corr	Dummy=1 if corruption reported as minor, moderate,major, or very severe obstacle to operations		
ic_crime	Dummy=1 if crime, theft, and disorder reported as minor, moderate,major, or very severe obstacle to operations		
ic_informal	Dummy=1 if practices of competitors in the informal sector reported as minor, moderate,major, or very severe obstacle to operations		

Table A.8: Secondary Bivariate Probit Model With Sample Selection - Further Estimations On Manufacturing (ES)

	telecoms2	electric2	water2	construction2	import2	tax2
Outcome Equation						
foreign	.	.	.	.	.	-4.527***
	.	.	.	.	.	(0.197)
security	-0.095	0.212*	0.274	-0.010	0.457***	0.204***
	(0.225)	(0.112)	(0.249)	(0.155)	(0.168)	(0.074)
age	-0.007	-0.000	0.004	-0.001	0.008	-0.000
	(0.013)	(0.007)	(0.014)	(0.008)	(0.010)	(0.004)
trade	0.363	0.221	0.731	0.541*	0.204	0.176
	(0.406)	(0.323)	(0.677)	(0.316)	(0.375)	(0.210)
ic_telecom	0.664**	0.139	0.344	0.097	0.003	0.030
	(0.289)	(0.116)	(0.247)	(0.120)	(0.162)	(0.075)
ic_elec	3.548***	0.332	-0.155	-0.281	-0.293	-0.489**
	(0.601)	(0.618)	(0.452)	(0.239)	(0.462)	(0.216)
ic_trans	-0.206	0.286	0.052	0.065	-0.028	0.351***
	(0.412)	(0.223)	(0.373)	(0.254)	(0.289)	(0.116)
ic_land	-0.029	0.099	0.179	-0.039	0.427**	-0.119
	(0.250)	(0.138)	(0.240)	(0.158)	(0.184)	(0.080)
ic_taxrates	0.939	0.199	0.837**	0.337	0.470*	0.422***
	(0.631)	(0.185)	(0.394)	(0.250)	(0.269)	(0.152)
ic_taxadmin	-0.355	-0.005	-0.961***	0.230	-0.267	-0.397***
	(0.496)	(0.169)	(0.340)	(0.260)	(0.245)	(0.112)
ic_customs	-0.181	0.056	-0.366	-0.042	-0.209	0.011
	(0.263)	(0.139)	(0.233)	(0.113)	(0.172)	(0.086)
ic_laborregs	0.017	0.075	0.198	0.251	-0.048	0.184**
	(0.265)	(0.134)	(0.222)	(0.162)	(0.169)	(0.089)
ic_educlabour	0.317	0.063	0.189	-0.124	0.146	-0.249***
	(0.238)	(0.127)	(0.220)	(0.155)	(0.166)	(0.081)
ic_buslic	0.658**	0.252*	0.425*	0.311**	0.301	0.328***
	(0.295)	(0.133)	(0.233)	(0.151)	(0.200)	(0.084)
ic_accfin	0.067	-0.173	0.129	0.134	-0.494*	-0.269**
	(0.311)	(0.166)	(0.301)	(0.168)	(0.253)	(0.112)
ic_costfin	-0.137	0.143	-0.244	-0.184	-0.087	0.062
	(0.329)	(0.177)	(0.337)	(0.248)	(0.263)	(0.123)
ic_polenviro	-0.212	-0.280**	-0.227	-0.220	-0.191	0.087
	(0.286)	(0.141)	(0.256)	(0.172)	(0.194)	(0.092)
ic_macroenviro	0.343	-0.152	0.029	-0.107	-0.101	-0.350***
	(0.342)	(0.157)	(0.318)	(0.198)	(0.199)	(0.092)
ic_corr	-0.360	-0.021	0.157	-0.114	0.661***	0.104
	(0.296)	(0.165)	(0.309)	(0.188)	(0.239)	(0.106)
ic_crime	0.180	0.481***	0.050	0.118	-0.021	0.287***
	(0.333)	(0.161)	(0.318)	(0.133)	(0.226)	(0.103)
ic_informal	0.133	-0.016	-0.037	-0.376	0.158	0.233**
	(0.349)	(0.147)	(0.278)	(0.000)	(0.217)	(0.094)
Constant	-5.894***	-1.815***	-0.988	1.726***	-1.085	-0.949***
	(0.767)	(0.664)	(0.864)	(0.326)	(0.794)	(0.258)
Selection Equation Ommitted						
Observations	2001	2001	2001	2001	2001	2001
Censored Observations	1759	1372	1759	1837	1599	332
Log-Likelihood	-716.0	-1425.9	-739.9	-580.2	-1076.6	-1615.6
Chi-Squared	849.6	40.8	29.7	.	37.1	1037.4
P Value	0.000	0.004	0.075	.	0.011	0.000
Rho	0.346	0.362	-0.105	-1.000	0.077	0.865
Wald (Rho=0)	0.42	4.24	0.11	12.99	0.03	5.02
Prob > chi2	0.5166	0.0394	0.7422	0.0003	0.8732	0.0250

$*p < 0.10, **p < 0.05, ***p < 0.01$

Dependent Variable in selection equation is Dummy=1 if firm requested public service, 0 otherwise (Table A.1). Dependent Variable in outcome equation is Dummy=1 if a bribe was requested/expected from the firm, 0 otherwise. (Huber-White) Heteroscedastic Robust standard errors in parenthesis. The base group for state-dummies is Abia state. The base group for sector dummies is “Other - Manufacturing”.

Table A.9: Secondary Bivariate Probit Model With Sample Selection - Further Estimations On All Firms (ES)

	telecoms2	electric2	water2	construction2	import2	tax2
Outcome Equation						
foreign	0.079	-0.234	-3.811***	-3.803**	-0.504	-0.744
	(0.804)	(0.716)	(0.463)	(1.778)	(0.794)	(0.499)
security	0.257**	0.283***	0.015	0.163	0.276**	0.131**
	(0.124)	(0.077)	(0.137)	(0.183)	(0.111)	(0.053)
age	0.001	0.007	-0.003	0.006	0.000	0.001
	(0.008)	(0.005)	(0.009)	(0.012)	(0.006)	(0.003)
trade	0.260	0.033	0.538	1.022	0.008	0.111
	(0.335)	(0.333)	(0.700)	(0.648)	(0.330)	(0.204)
ic_telecom	0.430***	0.266***	0.302**	0.067	0.117	0.048
	(0.138)	(0.079)	(0.139)	(0.149)	(0.105)	(0.054)
ic_elec	-0.648	-0.014	-0.109	-0.173	-0.632**	-0.348**
	(0.453)	(0.300)	(0.384)	(0.365)	(0.280)	(0.138)
ic_trans	-0.027	0.014	0.122	0.003	0.050	0.266***
	(0.186)	(0.119)	(0.185)	(0.240)	(0.160)	(0.078)
ic_land	0.031	-0.007	0.213	-0.097	0.167	-0.096*
	(0.129)	(0.089)	(0.146)	(0.194)	(0.114)	(0.056)
ic_taxrates	0.223	-0.014	-0.067	0.204	0.231	0.284***
	(0.231)	(0.118)	(0.187)	(0.244)	(0.158)	(0.100)
ic_taxadmin	-0.031	0.206*	0.035	0.363	-0.010	-0.169**
	(0.214)	(0.106)	(0.185)	(0.258)	(0.140)	(0.079)
ic_customs	0.017	-0.062	-0.225	0.173	-0.032	-0.069
	(0.126)	(0.083)	(0.137)	(0.179)	(0.105)	(0.055)
ic_laborregs	0.148	0.137	0.031	0.283	0.035	0.236***
	(0.151)	(0.090)	(0.148)	(0.212)	(0.115)	(0.063)
ic_educlabour	0.203	-0.009	0.030	-0.152	0.135	-0.188***
	(0.142)	(0.090)	(0.148)	(0.174)	(0.118)	(0.060)
ic_buslic	0.279*	0.228**	0.331**	0.639**	0.027	0.227***
	(0.150)	(0.091)	(0.150)	(0.323)	(0.134)	(0.063)
ic_accfin	0.339*	0.005	0.441*	0.494	-0.379**	-0.092
	(0.199)	(0.129)	(0.243)	(0.347)	(0.186)	(0.095)
ic_costfin	-0.329	0.227*	-0.208	-0.687	0.247	0.005
	(0.204)	(0.133)	(0.231)	(0.504)	(0.199)	(0.100)
ic_polenviro	-0.304*	-0.202**	-0.086	-0.308	-0.253**	0.077
	(0.166)	(0.095)	(0.170)	(0.204)	(0.122)	(0.067)
ic_macroenviro	-0.022	0.051	-0.059	-0.291	0.028	-0.180**
	(0.189)	(0.105)	(0.177)	(0.224)	(0.146)	(0.071)
ic_corr	-0.062	0.143	0.333*	0.194	0.448***	0.191**
	(0.175)	(0.115)	(0.201)	(0.212)	(0.164)	(0.078)
ic_crime	0.036	0.113	-0.123	-0.283	0.091	0.173**
	(0.182)	(0.111)	(0.180)	(0.296)	(0.158)	(0.077)
ic_informal	0.454**	0.177*	-0.128	-0.129	0.296**	0.272***
	(0.187)	(0.103)	(0.169)	(0.252)	(0.140)	(0.069)
Constant	-1.505**	-1.403***	-1.343**	0.455	-0.863**	-1.267***
	(0.596)	(0.358)	(0.580)	(1.596)	(0.404)	(0.165)
Selection Equation omitted						
Observations	3668	3668	3668	3668	3668	3668
Censored Observations	3031	2226	3128	3307	2794	706
Log-Likelihood	-1847.3	-3130.7	-1689.5	-1327.5	-2375.2	-3295.9
Chi-Squared	40.7	110.6	214.4	20.3	47.8	169.4
P Value	0.006	0.000	0.000	0.505	0.001	0.000
Rho	0.315	0.136	0.225	-0.427	0.140	0.920
Wald (Rho=0)	1.33	0.62	0.74	0.14	0.45	2.91
Prob>Chi2	0.2484	0.4301	0.3891	0.7083	0.5033	0.0881

$*p < 0.10, **p < 0.05, ***p < 0.01$

Dependent Variable in selection equation is Dummy=1 if firm requested public service, 0 otherwise. Dependent Variable in outcome equation is Dummy=1 if a bribe was requested/expected from the firm, 0 otherwise. (Huber-White) Heteroscedastic Robust standard errors in parenthesis. The base group for state-dummies is Abia state. The base group for sector dummies is "Rest Of Universe - Other".

Table A.10: Secondary Bivariate Probit Model With Sample Selection - Further Estimations On Manufacturing Companies (NBS)

	customs2	road2	goods_gov2	lic2	enviro2	res_work2	vehicle2	police2	traffic2
Outcome Equation									
foreign	0.378	0.576	4.327	0.390*	0.742***	2.561	0.192	7.202	0.153
	(0.377)	(0.393)	.	(0.208)	(0.260)	.	(2.559)	.	(0.127)
security	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.
age	-0.020	0.004	0.197***	0.000	-0.035***	-0.481	-0.001	-0.033***	0.002
	(0.021)	(0.010)	(0.024)	(0.000)	(0.008)	(0.000)	(0.054)	(0.011)	(0.008)
trade	.	-0.572*	7.229	-0.405***	-2.120***	-3.344	-0.007	-12.842	-0.244
	.	(0.341)	.	(0.153)	(0.446)	(0.000)	(14.712)	(0.000)	(0.158)
obs_tax_regs	-0.555***	0.014	3.985***	-0.511***	-1.718***	-0.142	-0.193	5.987	0.333**
	(0.196)	(0.331)	(0.368)	(0.177)	(0.384)	(0.000)	(2.575)	.	(0.133)
obs_infl	-1.127***	-0.021	-0.054	-0.168	-5.192	-0.551	-0.423	-9.151	-0.151
	(0.424)	(0.240)	(0.293)	(0.244)	(0.000)	(0.000)	(10.413)	(0.000)	(0.170)
obs_pol	0.537	-0.038	15.692	0.016	4.520***	0.172	0.009	7.774	0.312
	(0.356)	(0.198)	.	(0.183)	(0.348)	.	(6.717)	.	(0.195)
obs_change_law	-1.244***	-0.431	31.387	-0.105	1.651***	-10.019	-0.052	-16.589	-0.220*
	(0.460)	(0.351)	.	(0.156)	(0.382)	(0.000)	(9.801)	(0.000)	(0.128)
obs_crime	0.362	0.053	-6.093	0.440	-1.724	-11.477	0.287	6.437	0.160
	(0.427)	(0.514)	(0.000)	(0.400)	(0.000)	(0.000)	(0.308)	.	(0.106)
obs_corr	-0.729*	0.672	-33.677	-0.457	-0.759	16.254	0.673	-18.582***	-0.137
	(0.437)	(0.648)	(0.000)	(0.427)	(0.000)	.	(1.057)	(0.178)	(0.151)
obs_bus_reg	0.721***	-0.035	6.170	0.340	3.508***	-0.305	0.103	46.435	0.153**
	(0.222)	(0.212)	.	(0.439)	(0.297)	(0.000)	(1.390)	.	(0.060)
obs_unclear_laws	1.108**	0.247	-11.837***	-0.057	1.521	3.421	0.141	-25.227	-0.053
	(0.506)	(0.274)	(0.112)	(0.129)	.	.	(0.548)	(0.000)	(0.090)
Constant	0.865	-3.091***	-17.622	1.994***	-2.587***	4.628	-0.928	23.565	0.481
	(0.816)	(1.160)	(0.000)	(0.407)	(0.812)	.	(0.000)	.	(0.494)
Selection Equation									
Regional Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Constant	-5.836***	-5.546	-4.766***	-5.427***	-5.165***	-4.201***	-5.441	-4.402***	-5.498
	(1.565)	.	(0.074)	(0.116)	(0.089)	(0.097)	(17.371)	(0.102)	.
Pseudo R-squared									
Observations	314	309	325	314	309	320	309	320	315
Censored Observations	274	256	305	272	271	281	256	293	273
Log-Likelihood	-108.8	-143.9	-63.0	-117.3	-100.4	-99.9	-139.0	-78.7	-129.3
Chi-Squared	38.1	.	.	2634580.1	.	.	2147.8	.	.
P Value	0.000	.	.	0.000	.	.	0.000	.	.
Rho	1.000	1.000	-0.995	-1.000	-1.000	1.000	-1.000	-0.998	-1.000
Wald (Rho=0)	4.76	3.11	.	319.10	213.35	0.10	0.01	0.46	42.64
Prob>Chi2	0.0292	0.0776	.	0.0000	0.0000	0.7561	0.9371	0.4998	0.0000

$*p < 0.10, **p < 0.05, ***p < 0.01$

Dependent Variable in selection equation is Dummy=1 if firm met with public official, 0 otherwise (Table A.1). Dependent Variable in outcome equation is Dummy=1 if a bribe was paid to the official, 0 otherwise. The base group for state-dummies is Abia state.

Table A.11: Secondary Bivariate Probit Model With Sample Selection - Further Estimations On All Firms (NBS)

	customs2	road2	goods_gov2	lic2	goods_priv2	enviro2	res_work2	vehicle2	police2	traffic2	court2
	none	none	none	none	none	none	none	none	none	none	none
Outcome Equation											
foreign	-0.016	0.031	0.049	0.056	0.176	-0.151	0.010	-0.107	-0.274	-0.195	-0.424
	(0.584)	(0.258)	(0.060)	(0.595)	(0.307)	(0.341)	(0.157)	(0.195)	(0.182)	(0.151)	(1.896)
security	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.
age	-0.004	-0.003	0.000	-0.010	0.000	-0.004	-0.018*	-0.000	-0.004	-0.007*	-0.000
	(0.057)	(0.006)	(0.001)	(0.011)	(0.008)	(0.008)	(0.009)	(0.005)	(0.004)	(0.004)	(0.006)
trade	.	-0.062	-0.033	0.094	-0.244	-0.250	-0.433***	0.256*	-0.111	-0.205	-0.219
	.	(0.222)	(0.026)	(0.442)	(0.255)	(0.312)	(0.165)	(0.153)	(0.147)	(0.153)	(0.929)
obs_tax_regs	0.045	0.118	-0.032*	-0.029	0.200	0.087	0.021	0.079	0.169**	0.033	-0.000
	(1.434)	(0.120)	(0.017)	(0.267)	(0.167)	(0.143)	(0.092)	(0.087)	(0.072)	(0.088)	(1.317)
obs_infl	0.487	-0.167	0.269***	-0.046	0.307	0.175	-0.114	-0.071	0.044	-0.116	-0.266
	(2.929)	(0.171)	(0.044)	(0.290)	(0.211)	(0.200)	(0.108)	(0.110)	(0.149)	(0.099)	(0.837)
obs_pol	-0.152	0.263	0.036	0.163	-0.036	0.026	0.241**	0.140	0.079	0.184**	0.023
	(0.439)	(0.201)	(0.054)	(0.343)	(0.236)	(0.213)	(0.118)	(0.104)	(0.114)	(0.088)	(0.191)
obs_change_law	0.072	-0.074	0.202***	-0.061	-0.040	-0.032	-0.215	-0.066	-0.052	-0.033	-0.041
	(0.387)	(0.159)	(0.022)	(0.298)	(0.175)	(0.192)	(0.200)	(0.109)	(0.102)	(0.083)	(0.071)
obs_crime	-0.385	-0.072	-0.145***	-0.182	-0.427	-0.309	0.027	0.173	0.197	-0.099	0.238
	(0.638)	(0.234)	(0.010)	(0.290)	(0.327)	(0.287)	(0.169)	(0.170)	(0.181)	(0.153)	(0.426)
obs_corr	0.139	0.413*	0.211***	0.109	0.738**	0.175	0.293	0.230	-0.177	0.205*	0.205
	(1.920)	(0.216)	(0.027)	(0.191)	(0.346)	(0.322)	(0.224)	(0.146)	(0.160)	(0.105)	(1.290)
obs_bus_reg	-0.035	0.034	-0.055	0.221	0.202	0.377	0.001	-0.080	0.012	-0.052	0.158
	(0.111)	(0.156)	(0.057)	(0.396)	(0.231)	(0.277)	(0.088)	(0.095)	(0.080)	(0.086)	(0.784)
obs_unclear_laws	-0.053	-0.205	-0.127***	-0.125	-0.213	-0.197	-0.066	-0.020	0.021	0.030	-0.092
	(2.357)	(0.154)	(0.032)	(0.238)	(0.177)	(0.240)	(0.147)	(0.071)	(0.100)	(0.071)	(0.772)
Constant	-2.287	-2.478***	0.599***	-1.827	-4.277***	-1.966*	-2.375***	-3.156***	-2.415***	-1.975***	0.939
	(3.058)	(0.519)	(0.087)	(1.614)	(0.858)	(1.130)	(0.556)	(0.532)	(0.370)	(0.545)	(1.157)
Selection Equation omitted											
Observations	2035	2004	2045	1996	2044	2010	2035	1980	2025	2028	2050
Censored Observations	1935	1821	1950	1813	1942	1852	1917	1756	1877	1880	1964
Log-Likelihood	-374.2	-638.1	-388.3	-640.6	-406.3	-565.7	-432.5	-749.9	-572.9	-551.9	-334.9
Chi-Squared	.	7.6	18022.1	4.9	28.7	4.3	51.5	12.1	17.4	20.2	5791.6
P Value	.	0.750	0.000	0.937	0.003	0.960	0.000	0.353	0.096	0.042	0.000
Rho	1.000	0.671	-1.000	0.816	1.000	0.430	1.000	1.000	1.000	1.000	-1.000
Wald (Rho=0)	.	2.28	61.01	0.06	132.55	0.19	0.48	6.69	0.67	5.24	0.10
Prob>Chi2	.	0.1309	0.0000	0.8079	0.0000	0.6609	0.4868	0.0097	0.4131	0.0221	0.7562

$*p < 0.10, **p < 0.05, ***p < 0.01$

Dependent Variable in selection equation is Dummy=1 if firm met with public official, 0 otherwise (Table A.1). Dependent Variable in outcome equation is Dummy=1 if a bribe was paid to the official, 0 otherwise. The base group for state-dummies is Abia state. The base group for sector dummies is “Other community, social and personal services”.

A.1.2 Further Tables

Table A.12: Summary Statistics For NBS Data

VARIABLES	N	mean	sd	min	max
yrcomm	336	1,987	13.90	1,946	2,007
male	336	0.905	0.294	0	1
employee	336	2.045	1.180	1	4
obs_tax_regs	312	1.939	0.756	1	3
obs_infl	315	2.594	0.570	1	3
obs_pol	312	2.401	0.724	1	3
obs_change_law	294	2.252	0.747	1	3
obs_crime	322	2.699	0.606	1	3
obs_corr	315	2.667	0.628	1	3
obs_bus_reg	293	2	0.717	1	3
obs_unclear _laws	259	2.116	0.784	1	3
advantage _informal	239	2.318	1.065	1	4
bribe_offer _official	285	3.144	1.012	1	4
bribe_demand _official	283	3.237	0.981	1	4
bribe_knowledge _amount	256	3.367	0.940	1	4
bribe_offer _foreign_official	249	3.695	0.704	1	4
bribe_demand _foreign_official	246	3.744	0.629	1	4
corr_customs	272	3.195	1.029	1	4
corr_car	284	3.380	0.934	1	4
corr_buy_gov	261	3.368	0.966	1	4
corr_authorisation	265	3.264	1.014	1	4
corr_bus_permit	286	3.161	1.067	1	4
corr_buy_priv	246	3.443	0.878	1	4
corr_enviro_regs	281	3.231	1.025	1	4
corr_sanitary_regs	284	3.222	1.024	1	4
corr_hs_regs	262	3.401	0.941	1	4
corr_liti	226	3.619	0.746	1	4
corr_permit	248	3.403	0.956	1	4
corr_pub_proc	233	3.446	0.937	1	4
corr_trade_lic	237	3.451	0.922	1	4
corr_courts	238	3.622	0.768	1	4
bribe_dummy_direct	336	0.116	0.321	0	1

Table A.13: Summary Statistics For NBS Data, (cont'd)

VARIABLES	N	mean	sd	min	max
conda	336	0.176	0.381	0	1
condb	336	0.226	0.419	0	1
condc	336	0.0774	0.268	0	1
condd	336	0.185	0.388	0	1
conde	336	0.0863	0.281	0	1
condf	336	0.188	0.391	0	1
condg	336	0.152	0.359	0	1
condh	336	0.229	0.421	0	1
condi	336	0.122	0.328	0	1
condj	336	0.182	0.386	0	1
condk	336	0.0714	0.258	0	1
bribea	336	0.0536	0.226	0	1
bribeb	336	0.0446	0.207	0	1
bribec	336	0.0179	0.133	0	1
bribed	336	0.0476	0.213	0	1
bribee	336	0.00893	0.0942	0	1
bribef	336	0.0565	0.231	0	1
bribeg	336	0.0268	0.162	0	1
bribeh	336	0.0595	0.237	0	1
bribei	336	0.0565	0.231	0	1
bribej	336	0.0923	0.290	0	1
bribek	336	0.0208	0.143	0	1
bribe_naira	336	6,262	54,876	0	800,000
bribe_dummy _indirect	336	0.176	0.381	0	1
bribe_dummy	336	0.226	0.419	0	1
foreign	336	0.295	0.457	0	1
security	336	1	0	1	1
age	336	20.41	13.90	0	61
lage	335	2.686	0.945	0	4.111
trade	336	0.176	0.381	0	1

Table A.14: Correlation Matrix-NBS

	(1)																			
	conda	condb	condc	condd	conde	condf	condg	condh	condi	condj	condk	bribea	bribeb	bribec	bribed	bribee	bribef	bribeg	bribeh	
conda	1																			
condb	0.5***	1																		
condc	0.4***	0.4***	1																	
condd	0.4***	0.6***	0.3***	1																
conde	0.2***	0.4***	0.4***	0.3***	1															
condf	0.5***	0.6***	0.4***	0.6***	0.4***	1														
condg	0.5***	0.6***	0.4***	0.6***	0.3***	0.6***	1													
condh	0.5***	0.7***	0.4***	0.6***	0.4***	0.6***	0.5***	1												
condi	0.5***	0.5***	0.3***	0.5***	0.2***	0.5***	0.5***	0.4***	1											
condj	0.4***	0.5***	0.4***	0.4***	0.2***	0.4***	0.4***	0.4***	0.5***	1										
condk	0.3***	0.3***	0.2***	0.4***	0.4***	0.4***	0.4***	0.3***	0.5***	0.3***	1									
bribea	0.5***	0.09	0.03	0.2***	-0.07	0.3***	0.1*	0.2***	0.3***	0.3***	0.04	1								
bribeb	0.2***	0.4***	0.2***	0.2***	0.1*	0.2***	0.2***	0.4***	0.3***	0.4***	0.1*	0.3***	1							
bribec	0.1*	0.2***	0.5***	0.2**	0.2***	0.2**	0.2***	0.2***	0.2***	0.3***	0.1*	0.2**	0.5***	1						
bribed	0.2**	0.1**	0.04	0.5***	0.1*	0.3***	0.2**	0.2**	0.3***	0.3***	0.2**	0.4***	0.4***	0.08	1					
bribee	-0.04	0.2**	0.2***	0.1*	0.3***	0.1*	0.1*	0.2**	0.2**	0.2***	0.2***	-0.02	0.4***	0.5***	0.3***	1				
bribef	0.3***	0.2**	0.2**	0.3***	0.1*	0.5***	0.2***	0.3***	0.3***	0.3***	0.1*	0.5***	0.5***	0.4***	0.6***	0.3***	1			
bribeg	0.1*	0.2**	0.09	0.3***	0.2***	0.2***	0.4***	0.2***	0.2***	0.2***	0.10	0.3***	0.5***	0.1*	0.6***	0.4***	0.4***	1		
bribeh	0.2***	0.3***	0.1*	0.2***	0.06	0.3***	0.2**	0.5***	0.2***	0.4***	0.1*	0.4***	0.7***	0.3***	0.4***	0.4***	0.6***	0.4***	1	
bribei	0.3***	0.2**	0.2***	0.3***	0.1*	0.2***	0.2***	0.2***	0.7***	0.3***	0.2***	0.5***	0.4***	0.4***	0.4***	0.3***	0.4***	0.4***	0.4***	
bribej	0.2**	0.2***	0.2**	0.2***	0.09	0.2***	0.2***	0.3***	0.3***	0.7***	0.07	0.3***	0.6***	0.4***	0.4***	0.3***	0.4***	0.3***	0.6***	0.5
bribek	0.04	0.07	0.1*	0.3***	0.1	0.1**	0.1*	0.2**	0.3***	0.1**	0.5***	0.2**	0.3***	0.3***	0.4***	0.4***	0.3***	0.2***	0.3***	0.4
bribe_naira	0.06	-0.008	-0.02	0.2***	-0.03	0.1*	-0.001	0.05	0.1*	0.09	0.2***	0.2**	0.05	-0.02	0.4***	-0.009	0.2***	0.07	0.2**	0.2
bribe_dummy	0.2***	0.2***	0.2**	0.2***	0.09	0.3***	0.1*	0.3***	0.3***	0.4***	0.10	0.4***	0.4***	0.2***	0.4***	0.2**	0.5***	0.3***	0.5***	0.5

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.15: Bribery, by Region

region	Dummy=1 if bribe paid, 0 otherwise 0	1	Total	Observations
abia Freq	28.0	26.0	54.0	54
Cell%	1.4	1.3	2.7	
Row%	51.9	48.1	100.0	
Col%	2.9	2.5	2.7	
adamawa Freq	8.0	28.0	36.0	36
Cell%	0.4	1.4	1.8	
Row%	22.2	77.8	100.0	
Col%	0.8	2.7	1.8	
akwa ibom Freq	9.0	48.0	57.0	57
Cell%	0.4	2.4	2.8	
Row%	15.8	84.2	100.0	
Col%	0.9	4.7	2.8	
anambra Freq	28.0	25.0	53.0	53
Cell%	1.4	1.2	2.6	
Row%	52.8	47.2	100.0	
Col%	2.9	2.4	2.6	
bauchi Freq	42.0	18.0	60.0	60
Cell%	2.1	0.9	3.0	
Row%	70.0	30.0	100.0	
Col%	4.3	1.7	3.0	
bayelsa Freq	4.0	6.0	10.0	10
Cell%	0.2	0.3	0.5	
Row%	40.0	60.0	100.0	
Col%	0.4	0.6	0.5	
benue Freq	0.0	60.0	60.0	60
Cell%	0.0	3.0	3.0	
Row%	0.0	100.0	100.0	
Col%	0.0	5.8	3.0	
borno Freq	26.0	20.0	46.0	46
Cell%	1.3	1.0	2.3	
Row%	56.5	43.5	100.0	
Col%	2.7	1.9	2.3	
cross river Freq	33.0	12.0	45.0	45
Cell%	1.6	0.6	2.2	
Row%	73.3	26.7	100.0	
Col%	3.4	1.2	2.2	
delta Freq	28.0	12.0	40.0	40
Cell%	1.4	0.6	2.0	
Row%	70.0	30.0	100.0	
Col%	2.9	1.2	2.0	
ebonyi Freq	27.0	38.0	65.0	65
Cell%	1.3	1.9	3.2	
Row%	41.5	58.5	100.0	
Col%	2.8	3.7	3.2	
edo Freq	20.0	18.0	38.0	38
Cell%	1.0	0.9	1.9	
Row%	52.6	47.4	100.0	
Col%	2.1	1.7	1.9	
ekiti Freq	18.0	44.0	62.0	62
Cell%	0.9	2.2	3.1	
Row%	29.0	71.0	100.0	
Col%	1.9	4.3	3.1	
enugu Freq	29.0	25.0	54.0	54
Cell%	1.4	1.2	2.7	
Row%	53.7	46.3	100.0	
Col%	3.0	2.4	2.7	
gombe Freq	39.0	23.0	62.0	62
Cell%	1.9	1.1	3.1	
Row%	62.9	37.1	100.0	
Col%	4.0	2.2	3.1	
imo Freq	4.0	34.0	38.0	38
Cell%	0.2	1.7	1.9	
Row%	10.5	89.5	100.0	
Col%	0.4	3.3	1.9	

Table A.16: Bribery, by Region, (cont'd)

region	Dummy=1 if bribe paid, 0 otherwise 0	1	Total	Obs.
jigawa Freq	37.0	27.0	64.0	64
Cell%	1.8	1.3	3.2	
Row%	57.8	42.2	100.0	
Col%	3.8	2.6	3.2	
kaduna Freq	54.0	22.0	76.0	76
Cell%	2.7	1.1	3.8	
Row%	71.1	28.9	100.0	
Col%	5.6	2.1	3.8	
kano Freq	60.0	35.0	95.0	95
Cell%	3.0	1.7	4.7	
Row%	63.2	36.8	100.0	
Col%	6.2	3.4	4.7	
katsina Freq	44.0	17.0	61.0	61
Cell%	2.2	0.8	3.0	
Row%	72.1	27.9	100.0	
Col%	4.5	1.7	3.0	
kebbi Freq	34.0	32.0	66.0	66
Cell%	1.7	1.6	3.3	
Row%	51.5	48.5	100.0	
Col%	3.5	3.1	3.3	
kogi Freq	1.0	32.0	33.0	33
Cell%	0.0	1.6	1.6	
Row%	3.0	97.0	100.0	
Col%	0.1	3.1	1.6	
kwara Freq	22.0	27.0	49.0	49
Cell%	1.1	1.3	2.4	
Row%	44.9	55.1	100.0	
Col%	2.3	2.6	2.4	
lagos Freq	89.0	21.0	110.0	110
Cell%	4.4	1.0	5.5	
Row%	80.9	19.1	100.0	
Col%	9.2	2.0	5.5	
nassarawa Freq	1.0	60.0	61.0	61
Cell%	0.0	3.0	3.0	
Row%	1.6	98.4	100.0	
Col%	0.1	5.8	3.0	
niger Freq	38.0	38.0	76.0	76
Cell%	1.9	1.9	3.8	
Row%	50.0	50.0	100.0	
Col%	3.9	3.7	3.8	
ogun Freq	50.0	30.0	80.0	80
Cell%	2.5	1.5	4.0	
Row%	62.5	37.5	100.0	
Col%	5.1	2.9	4.0	
ondo Freq	27.0	24.0	51.0	51
Cell%	1.3	1.2	2.5	
Row%	52.9	47.1	100.0	
Col%	2.8	2.3	2.5	
osun Freq	34.0	21.0	55.0	55
Cell%	1.7	1.0	2.7	
Row%	61.8	38.2	100.0	
Col%	3.5	2.0	2.7	
oyo Freq	18.0	12.0	30.0	30
Cell%	0.9	0.6	1.5	
Row%	60.0	40.0	100.0	
Col%	1.9	1.2	1.5	
plateau Freq	27.0	24.0	51.0	51
Cell%	1.3	1.2	2.5	
Row%	52.9	47.1	100.0	
Col%	2.8	2.3	2.5	

Table A.17: Bribery, by Region, (cont'd)

	Dummy=1 if bribe paid, 0 otherwise			
region	**0**	**1**	**Total**	**Observations**
rivers Freq	7.0	25.0	32.0	32
Cell%	0.3	1.2	1.6	
Row%	21.9	78.1	100.0	
Col%	0.7	2.4	1.6	
sokoto Freq	21.0	19.0	40.0	40
Cell%	1.0	0.9	2.0	
Row%	52.5	47.5	100.0	
Col%	2.2	1.8	2.0	
taraba Freq	0.0	35.0	35.0	35
Cell%	0.0	1.7	1.7	
Row%	0.0	100.0	100.0	
Col%	0.0	3.4	1.7	
yobe Freq	0.0	15.0	15.0	15
Cell%	0.0	0.7	0.7	
Row%	0.0	100.0	100.0	
Col%	0.0	1.5	0.7	
zamfara Freq	27.0	47.0	74.0	74
Cell%	1.3	2.3	3.7	
Row%	36.5	63.5	100.0	
Col%	2.8	4.6	3.7	
FCT (abuja) Freq	38.0	29.0	67.0	67
Cell%	1.9	1.4	3.3	
Row%	56.7	43.3	100.0	
Col%	3.9	2.8	3.3	
Total Freq	972.0	1029.0	2001.0	2,001
Cell%	48.6	51.4	100.0	
Row%	48.6	51.4	100.0	
Col%	100.0	100.0	100.0	

Table A.18: Bribery, by Industry

industry	Dummy=1 if bribe paid, 0 otherwise 0	1	Total	Observations
mfg - food Freq	208.0	165.0	373.0	373
Cell%	10.4	8.2	18.6	
Row%	55.8	44.2	100.0	
Col%	21.4	16.0	18.6	
mfg - garments Freq	197.0	148.0	345.0	345
Cell%	9.8	7.4	17.2	
Row%	57.1	42.9	100.0	
Col%	20.3	14.4	17.2	
mfg - textiles Freq	11.0	10.0	21.0	21
Cell%	0.5	0.5	1.0	
Row%	52.4	47.6	100.0	
Col%	1.1	1.0	1.0	
mfg - machinery & equipment Freq	9.0	7.0	16.0	16
Cell%	0.4	0.3	0.8	
Row%	56.3	43.8	100.0	
Col%	0.9	0.7	0.8	
mfg - chemicals Freq	9.0	12.0	21.0	21
Cell%	0.4	0.6	1.0	
Row%	42.9	57.1	100.0	
Col%	0.9	1.2	1.0	
mfg - electronics Freq	5.0	1.0	6.0	6
Cell%	0.2	0.0	0.3	
Row%	83.3	16.7	100.0	
Col%	0.5	0.1	0.3	
mfg - non-metallic minerals Freq	89.0	106.0	195.0	195
Cell%	4.4	5.3	9.7	
Row%	45.6	54.4	100.0	
Col%	9.2	10.3	9.7	
mfg - wood, wood products & furniture Freq	210.0	264.0	474.0	474
Cell%	10.5	13.2	23.7	
Row%	44.3	55.7	100.0	
Col%	21.6	25.7	23.7	
mfg - metal & metal products Freq	149.0	147.0	296.0	296
Cell%	7.4	7.3	14.8	
Row%	50.3	49.7	100.0	
Col%	15.3	14.3	14.8	
mfg - other manufacturing Freq	85.0	169.0	254.0	254
Cell%	4.2	8.4	12.7	
Row%	33.5	66.5	100.0	
Col%	8.7	16.4	12.7	
Total Freq	972.0	1029.0	2001.0	2,001
Cell%	48.6	51.4	100.0	
Row%	48.6	51.4	100.0	
Col%	100.0	100.0	100.0	

Table A.19: Bribery, by Industry, with 95% Confidence Intervals

industry	**Dummy=1 if bribe paid, 0 otherwise** **0**			**1**			**Total**	**Observations**
	Row%	(Std.Err.)	95% C.I	Row%	(Std.Err.)	95% C.I	Row%	
mfg - food	61.9	(8.4)	[44.6,76.7]	38.1	(8.4)	[23.3,55.4]	100.0	373
mfg - garments	57.5	(3.6)	[50.3,64.4]	42.5	(3.6)	[35.6,49.7]	100.0	345
mfg - textiles	34.7	(12.7)	[15.1,61.4]	65.3	(12.7)	[38.6,84.9]	100.0	21
mfg - machinery & equipment	46.5	(17.9)	[17.5,78.1]	53.5	(17.9)	[21.9,82.5]	100.0	16
mfg - chemicals	36.8	(11.3)	[18.2,60.2]	63.2	(11.3)	[39.8,81.8]	100.0	21
mfg - electronics	89.7	(10.4)	[48.8,98.7]	10.3	(10.4)	[1.3,51.2]	100.0	6
mfg - non-metallic minerals	55.0	(5.7)	[43.8,65.6]	45.0	(5.7)	[34.4,56.2]	100.0	195
mfg - wood, wood products & furniture	47.3	(4.1)	[39.5,55.3]	52.7	(4.1)	[44.7,60.5]	100.0	474
mfg - metal & metal products	56.4	(6.2)	[44.0,68.0]	43.6	(6.2)	[32.0,56.0]	100.0	296
mfg - other manufacturing	32.5	(5.4)	[22.8,43.9]	67.5	(5.4)	[56.1,77.2]	100.0	254
Total	51.6	(2.5)	[46.7,56.4]	48.4	(2.5)	[43.6,53.3]	100.0	2,001

Table A.20: Bribery Reporting Behaviour, by Region

	0=no bribe, 1=sales,2=ngn,3=missing				
region	**No Bribe**	**Sales Reporter**	**Naira Reporter**	**Total**	**Observations**
abia Freq	28.0	18.0	8.0	54.0	54
Cell%	1.4	0.9	0.4	2.7	
Row%	51.9	33.3	14.8	100.0	
Col%	2.9	2.2	3.5	2.7	
adamawa Freq	8.0	19.0	9.0	36.0	36
Cell%	0.4	0.9	0.4	1.8	
Row%	22.2	52.8	25.0	100.0	
Col%	0.8	2.4	4.0	1.8	
akwa ibom Freq	9.0	22.0	26.0	57.0	57
Cell%	0.4	1.1	1.3	2.8	
Row%	15.8	38.6	45.6	100.0	
Col%	0.9	2.7	11.5	2.8	
anambra Freq	28.0	14.0	11.0	53.0	53
Cell%	1.4	0.7	0.5	2.6	
Row%	52.8	26.4	20.8	100.0	
Col%	2.9	1.7	4.8	2.6	
bauchi Freq	42.0	18.0	0.0	60.0	60
Cell%	2.1	0.9	0.0	3.0	
Row%	70.0	30.0	0.0	100.0	
Col%	4.3	2.2	0.0	3.0	
bayelsa Freq	4.0	6.0	0.0	10.0	10
Cell%	0.2	0.3	0.0	0.5	
Row%	40.0	60.0	0.0	100.0	
Col%	0.4	0.7	0.0	0.5	
benue Freq	0.0	38.0	22.0	60.0	60
Cell%	0.0	1.9	1.1	3.0	
Row%	0.0	63.3	36.7	100.0	
Col%	0.0	4.7	9.7	3.0	
borno Freq	26.0	18.0	2.0	46.0	46
Cell%	1.3	0.9	0.1	2.3	
Row%	56.5	39.1	4.3	100.0	
Col%	2.7	2.2	0.9	2.3	
cross river Freq	33.0	12.0	0.0	45.0	45
Cell%	1.6	0.6	0.0	2.2	
Row%	73.3	26.7	0.0	100.0	
Col%	3.4	1.5	0.0	2.2	
delta Freq	28.0	1.0	11.0	40.0	40
Cell%	1.4	0.0	0.5	2.0	
Row%	70.0	2.5	27.5	100.0	
Col%	2.9	0.1	4.8	2.0	
ebonyi Freq	27.0	31.0	7.0	65.0	65
Cell%	1.3	1.5	0.3	3.2	
Row%	41.5	47.7	10.8	100.0	
Col%	2.8	3.9	3.1	3.2	
edo Freq	20.0	11.0	7.0	38.0	38
Cell%	1.0	0.5	0.3	1.9	
Row%	52.6	28.9	18.4	100.0	
Col%	2.1	1.4	3.1	1.9	
ekiti Freq	18.0	41.0	3.0	62.0	62
Cell%	0.9	2.0	0.1	3.1	
Row%	29.0	66.1	4.8	100.0	
Col%	1.9	5.1	1.3	3.1	
enugu Freq	29.0	17.0	8.0	54.0	54
Cell%	1.4	0.8	0.4	2.7	
Row%	53.7	31.5	14.8	100.0	
Col%	3.0	2.1	3.5	2.7	
gombe Freq	39.0	19.0	4.0	62.0	62
Cell%	1.9	0.9	0.2	3.1	
Row%	62.9	30.6	6.5	100.0	
Col%	4.0	2.4	1.8	3.1	
imo Freq	4.0	18.0	16.0	38.0	38
Cell%	0.2	0.9	0.8	1.9	
Row%	10.5	47.4	42.1	100.0	
Col%	0.4	2.2	7.0	1.9	

Table A.21: Bribery Reporting Behaviour, by Region (cont'd)

	0=no bribe, 1=sales,2=ngn,3=missing				
region	**No Bribe**	**Sales Reporter**	**Naira Reporter**	**Total**	**Observations**
jigawa Freq	37.0	22.0	5.0	64.0	64
Cell%	1.8	1.1	0.2	3.2	
Row%	57.8	34.4	7.8	100.0	
Col%	3.8	2.7	2.2	3.2	
kaduna Freq	54.0	17.0	5.0	76.0	76
Cell%	2.7	0.8	0.2	3.8	
Row%	71.1	22.4	6.6	100.0	
Col%	5.6	2.1	2.2	3.8	
kano Freq	60.0	33.0	2.0	95.0	95
Cell%	3.0	1.6	0.1	4.7	
Row%	63.2	34.7	2.1	100.0	
Col%	6.2	4.1	0.9	4.7	
katsina Freq	44.0	11.0	6.0	61.0	61
Cell%	2.2	0.5	0.3	3.0	
Row%	72.1	18.0	9.8	100.0	
Col%	4.5	1.4	2.6	3.0	
kebbi Freq	34.0	28.0	4.0	66.0	66
Cell%	1.7	1.4	0.2	3.3	
Row%	51.5	42.4	6.1	100.0	
Col%	3.5	3.5	1.8	3.3	
kogi Freq	1.0	32.0	0.0	33.0	33
Cell%	0.0	1.6	0.0	1.6	
Row%	3.0	97.0	0.0	100.0	
Col%	0.1	4.0	0.0	1.6	
kwara Freq	22.0	27.0	0.0	49.0	49
Cell%	1.1	1.3	0.0	2.4	
Row%	44.9	55.1	0.0	100.0	
Col%	2.3	3.4	0.0	2.4	
lagos Freq	89.0	17.0	4.0	110.0	110
Cell%	4.4	0.8	0.2	5.5	
Row%	80.9	15.5	3.6	100.0	
Col%	9.2	2.1	1.8	5.5	
nassarawa Freq	1.0	60.0	0.0	61.0	61
Cell%	0.0	3.0	0.0	3.0	
Row%	1.6	98.4	0.0	100.0	
Col%	0.1	7.5	0.0	3.0	
niger Freq	38.0	33.0	5.0	76.0	76
Cell%	1.9	1.6	0.2	3.8	
Row%	50.0	43.4	6.6	100.0	
Col%	3.9	4.1	2.2	3.8	
ogun Freq	50.0	28.0	2.0	80.0	80
Cell%	2.5	1.4	0.1	4.0	
Row%	62.5	35.0	2.5	100.0	
Col%	5.1	3.5	0.9	4.0	
ondo Freq	27.0	21.0	3.0	51.0	51
Cell%	1.3	1.0	0.1	2.5	
Row%	52.9	41.2	5.9	100.0	
Col%	2.8	2.6	1.3	2.5	
osun Freq	34.0	20.0	1.0	55.0	55
Cell%	1.7	1.0	0.0	2.7	
Row%	61.8	36.4	1.8	100.0	
Col%	3.5	2.5	0.4	2.7	
oyo Freq	18.0	11.0	1.0	30.0	30
Cell%	0.9	0.5	0.0	1.5	
Row%	60.0	36.7	3.3	100.0	
Col%	1.9	1.4	0.4	1.5	
plateau Freq	27.0	22.0	2.0	51.0	51
Cell%	1.3	1.1	0.1	2.5	
Row%	52.9	43.1	3.9	100.0	
Col%	2.8	2.7	0.9	2.5	
rivers Freq	7.0	1.0	24.0	32.0	32
Cell%	0.3	0.0	1.2	1.6	
Row%	21.9	3.1	75.0	100.0	
Col%	0.7	0.1	10.6	1.6	

Table A.22: Bribery Reporting Behaviour, by Region (cont'd)

	0=no bribe, 1=sales,2=ngn,3=missing				
region	**No Bribe**	**Sales Reporter**	**Naira Reporter**	**Total**	**Observations**
sokoto Freq	21.0	13.0	6.0	40.0	40
Cell%	1.0	0.6	0.3	2.0	
Row%	52.5	32.5	15.0	100.0	
Col%	2.2	1.6	2.6	2.0	
taraba Freq	0.0	17.0	18.0	35.0	35
Cell%	0.0	0.8	0.9	1.7	
Row%	0.0	48.6	51.4	100.0	
Col%	0.0	2.1	7.9	1.7	
yobe Freq	0.0	15.0	0.0	15.0	15
Cell%	0.0	0.7	0.0	0.7	
Row%	0.0	100.0	0.0	100.0	
Col%	0.0	1.9	0.0	0.7	
zamfara Freq	27.0	47.0	0.0	74.0	74
Cell%	1.3	2.3	0.0	3.7	
Row%	36.5	63.5	0.0	100.0	
Col%	2.8	5.9	0.0	3.7	
FCT (abuja) Freq	38.0	24.0	5.0	67.0	67
Cell%	1.9	1.2	0.2	3.3	
Row%	56.7	35.8	7.5	100.0	
Col%	3.9	3.0	2.2	3.3	
Total Freq	972.0	802.0	227.0	2001.0	2,001
Cell%	48.6	40.1	11.3	100.0	
Row%	48.6	40.1	11.3	100.0	
Col%	100.0	100.0	100.0	100.0	

Table A.23: Correlation Matrix - WB

	(1)					
	eprofit	university	foreign	age	security_naira	ind_loc_eprofit
eprofit	1					
university	0.05*	1				
foreign	0.01	0.05*	1			
age	-0.008	0.03	-0.004	1		
security_naira	0.2***	0.1***	0.01	-0.03	1	
ind_loc_eprofit	0.2***	0.07**	0.002	-0.04	0.1***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.24: Probit Estimations On The Propensity To Bribe

	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12
infraserv	0.175***			0.170***	0.175***	0.171***	0.192***	0.180***	0.182***	0.175***	0.177***	0.172***
	(0.057)			(0.057)	(0.058)	(0.058)	(0.058)	(0.059)	(0.059)	(0.059)	(0.059)	(0.067)
trade		0.039		0.023	0.029	-0.001	0.021	-0.011	-0.006	-0.008	0.005	-0.000
		(0.217)		(0.216)	(0.216)	(0.216)	(0.217)	(0.216)	(0.216)	(0.216)	(0.216)	(0.241)
tax_percentage			0.006***	0.006***	0.006***	0.006***	0.005***	0.006***	0.006***	0.006***	0.006***	0.004***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
profit					-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000*
					(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
capital_labour						0.000	0.000	0.000	0.000	0.000	0.000	0.000
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
un_sunk_cost							-4.503***	-4.341***	-4.234***	-4.377***	-4.323***	0.060
							(1.121)	(1.124)	(1.127)	(1.130)	(1.131)	(1.319)
competitors_2								-0.242	-0.242	-0.245	-0.234	0.051
								(0.288)	(0.288)	(0.287)	(0.288)	(0.340)
competitors_3								-0.251	-0.253	-0.239	-0.228	0.283
								(0.181)	(0.181)	(0.181)	(0.181)	(0.225)
competitors_4								-0.468***	-0.476***	-0.461***	-0.454***	0.146
								(0.172)	(0.172)	(0.172)	(0.173)	(0.217)
gov_customer									0.526**	0.515**	0.509*	0.320
									(0.259)	(0.260)	(0.261)	(0.294)
regulation_realtime										0.016**	0.018**	0.032***
										(0.008)	(0.008)	(0.008)
external_consultant_naira											-0.000**	0.000
											(0.000)	(0.000)
african												-0.499***
												(0.157)
Constant	-0.014	0.035	-0.356***	-0.401***	-0.392***	-0.410***	-4.387***	-3.852***	-3.762***	-3.937***	-3.910***	2.600**
	(0.032)	(0.028)	(0.078)	(0.079)	(0.081)	(0.081)	(0.995)	(1.013)	(1.015)	(1.020)	(1.021)	(1.286)
Industry Dummies												YES
Region Dummies												YES
Pseudo R-squared	0.003	0.000	0.011	0.014	0.014	0.016	0.022	0.028	0.029	0.031	0.033	0.200
Observations	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001
Log-Likelihood	-1381.4	-1386.2	-1371.2	-1366.7	-1366.6	-1363.6	-1355.4	-1347.4	-1345.4	-1343.1	-1341.1	-1109.2
Chi-Squared	9.3	0.0	29.4	38.2	38.7	40.9	56.5	69.3	73.6	78.6	82.3	348.4
Ind-Chi-Squared												9.4
Ind-P Value												0.224
Reg-Chi-Squared												193
Reg-P Value												0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$

Dependent variable is a dummy=1 if firm reported a positive bribe payment; 0 otherwise.

Table A.25: Marginal And Impact Effects From First Stage Probit Models

	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12
infraserv	0.070***			0.068***	0.070***	0.068***	0.077***	0.072***	0.073***	0.070***	0.071***	0.068***
	(0.023)			(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.024)	(0.026)
trade (d)		0.015		0.009	0.012	-0.000	0.008	-0.005	-0.002	-0.003	0.002	-0.000
		(0.086)		(0.086)	(0.086)	(0.086)	(0.086)	(0.086)	(0.086)	(0.086)	(0.086)	(0.095)
tax_percentage			0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.001***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
profit					-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000*
					(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
capital_labour						0.000	0.000	0.000	0.000	0.000	0.000	0.000
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
un_sunk_cost							-1.795***	-1.730***	-1.688***	-1.745***	-1.723***	0.024
							(0.447)	(0.448)	(0.449)	(0.451)	(0.451)	(0.520)
competitors_2 (d)								-0.096	-0.096	-0.097	-0.093	0.020
								(0.113)	(0.113)	(0.112)	(0.113)	(0.133)
competitors_3 (d)								-0.100	-0.100	-0.095	-0.091	0.109
								(0.071)	(0.071)	(0.071)	(0.072)	(0.084)
competitors_4 (d)								-0.183***	-0.186***	-0.180***	-0.177***	0.058
								(0.065)	(0.064)	(0.065)	(0.065)	(0.086)
gov_customer (d)									0.199**	0.195**	0.193**	0.121
									(0.089)	(0.089)	(0.090)	(0.105)
regulation_realtime										0.006**	0.007**	0.013***
										(0.003)	(0.003)	(0.003)
external_consultant_naira											-0.000**	0.000
											(0.000)	(0.000)
african (d)												-0.183***
												(0.052)

(d) for discrete change of dummy variable from 0 to 1.

$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A.26: Regressions On The Amount Of Bribe

	m0	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10
bribereport	-42.928***	-42.958***	-38.487***	-38.028***	-37.815***	-37.711***	-37.993***	-38.123***	-36.998***	-37.048***	-35.826***
	(2.401)	(2.403)	(2.484)	(2.473)	(2.454)	(2.476)	(2.511)	(2.515)	(2.589)	(2.601)	(3.065)
infraserv	1.999	1.955	-4.375	-4.367	-2.880	-3.002	-2.839	-2.511	-2.311	-2.314	2.102
	(3.680)	(3.675)	(3.487)	(3.471)	(3.392)	(3.420)	(3.402)	(3.390)	(3.306)	(3.307)	(3.653)
import	-9.195										
	(11.654)										
export	4.416										
	(19.624)										
tax_percentage	0.024	0.023	0.026	0.020	0.030	0.032	0.037	0.034	0.026	0.026	-0.058
	(0.065)	(0.065)	(0.051)	(0.051)	(0.050)	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)	(0.062)
trade		-5.737	-12.855	-16.929**	-14.821*	-14.916*	-14.433*	-14.290*	-13.834*	-13.831*	-9.213
		(10.454)	(7.899)	(7.905)	(8.361)	(8.333)	(8.343)	(8.316)	(7.792)	(7.826)	(7.084)
eprofit			0.177***	0.174***	0.172***	0.172***	0.172***	0.172***	0.165***	0.165***	0.155***
			(0.040)	(0.040)	(0.039)	(0.039)	(0.039)	(0.039)	(0.035)	(0.035)	(0.035)
capital_labour				0.006	0.006	0.006	0.005	0.005	0.006	0.006	0.006**
				(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
eun_sunk_cost					-113.584***	-113.434***	-115.967***	-115.515***	-122.198***	-121.736***	
					(32.891)	(33.463)	(33.461)	(33.388)	(35.115)	(35.253)	
competitors_2						3.991	4.007	4.692	3.972	3.909	-2.171
						(8.271)	(8.272)	(8.237)	(8.196)	(8.194)	(10.315)
competitors_3						0.863	0.666	0.385	-0.847	-1.086	3.922
						(4.628)	(4.631)	(4.652)	(4.772)	(4.763)	(6.051)
competitors_4						0.108	-0.331	-0.521	-1.390	-1.622	7.921
						(4.165)	(4.160)	(4.188)	(4.258)	(4.287)	(6.716)
gov_customer							22.518	22.754	23.704*	23.716*	18.169
							(13.955)	(13.959)	(14.258)	(14.258)	(13.760)
regulation_realtime								-0.482*	-0.799**	-0.797**	0.073
								(0.248)	(0.368)	(0.369)	(0.469)
eexternal_consultant_naira									0.551	0.552	0.650
									(0.459)	(0.460)	(0.488)
african										1.659	4.340
										(3.866)	(3.962)
un_sunk_cost											-5.058
											(59.839)
Industry Dummies											YES
Region Dummies											YES
Constant	47.195***	47.264***	19.600***	18.938***	8.644	8.203	7.625	9.312	10.685	9.388	44.830
	(5.431)	(5.430)	(6.039)	(5.984)	(7.069)	(8.160)	(8.157)	(8.138)	(7.464)	(7.601)	(48.844)
F	65.652	82.062	53.193	44.745	39.081	27.972	25.174	23.123	23.272	21.640	.
Adjusted R-squared	0.075	0.076	0.456	0.461	0.467	0.465	0.467	0.467	0.476	0.476	0.527
Observations	1029	1029	1029	1029	1029	1029	1029	1029	1029	1029	1029
Ind-F											1.6
Ind-P Value											0.128
Reg-F											4
Reg-P Value											0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$

Dependent variable is bribe amount in '000 Naira.

Table A.27: Tobit Estimations On Bribe Amounts

	m0	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10
model											
infraserv	10.027***	10.037***	5.222	5.010	6.846*	6.035*	6.260*	6.014*	6.013*	6.157*	6.373*
	(3.861)	(3.859)	(3.588)	(3.571)	(3.560)	(3.573)	(3.551)	(3.543)	(3.544)	(3.528)	(3.429)
import	-4.338										
	(13.769)										
export	-12.407										
	(24.374)										
trade		-9.123	-17.219	-21.026*	-18.769	-20.928*	-20.357*	-20.423*	-19.876*	-20.854*	-10.933
		(12.543)	(12.580)	(11.970)	(12.028)	(11.778)	(11.754)	(11.766)	(11.445)	(11.515)	(11.756)
tax_percentage	0.244***	0.245***	0.225***	0.218***	0.225***	0.242***	0.248***	0.247***	0.248***	0.252***	0.065
	(0.068)	(0.068)	(0.061)	(0.061)	(0.061)	(0.062)	(0.062)	(0.062)	(0.061)	(0.061)	(0.069)
eprofit			0.116***	0.112***	0.110***	0.111***	0.110***	0.110***	0.111***	0.112***	0.106***
			(0.032)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.030)	(0.030)	(0.028)
capital_labour				0.009**	0.009**	0.009**	0.009**	0.009**	0.009**	0.009**	0.007***
				(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
eun_sunk_cost					-142.581***	-131.359***	-134.836***	-136.811***	-136.586***	-140.175***	
					(41.695)	(41.813)	(41.912)	(41.955)	(42.166)	(42.395)	
competitors_2						-7.823	-7.796	-8.063	-7.866	-7.484	1.887
						(13.880)	(13.875)	(13.841)	(13.819)	(13.778)	(11.910)
competitors_3						-8.973	-9.131	-8.785	-8.654	-6.673	18.610**
						(8.047)	(8.047)	(8.028)	(8.084)	(8.052)	(8.359)
competitors_4						-22.119***	-22.658***	-22.284***	-22.174***	-20.148***	13.238
						(7.483)	(7.482)	(7.468)	(7.519)	(7.490)	(8.665)
gov_customer							34.672*	34.404*	34.258*	34.264*	20.643
							(17.769)	(17.776)	(17.768)	(17.807)	(16.180)
regulation_realtime								0.452	0.475	0.460	1.522***
								(0.340)	(0.354)	(0.350)	(0.366)
eexternal_consultant_naira									-0.051	-0.051	0.221
									(0.221)	(0.221)	(0.240)
african										-17.734***	-13.315**
										(6.113)	(6.237)
un_sunk_cost											14.372
											(66.191)
Constant	-31.799***	-31.869***	-44.133***	-44.854***	-57.080***	-38.534***	-39.181***	-40.780***	-41.020***	-26.648**	61.734
	(7.603)	(7.615)	(9.548)	(9.580)	(11.642)	(12.899)	(12.929)	(13.084)	(12.802)	(12.311)	(56.849)
sigma	77.757***	77.752***	69.983***	69.556***	69.172***	68.966***	68.859***	68.863***	68.870***	68.809***	59.710***
	(10.586)	(10.586)	(7.057)	(7.029)	(6.942)	(6.990)	(7.006)	(7.006)	(7.007)	(6.998)	(6.396)
Industry Dummies											YES
Region Dummies											YES
Pseudo R-squared	0.0015	0.001	0.017	0.018	0.019	0.020	0.020	0.020	0.020	0.021	0.057
Observations	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001
Log-Likelihood	-6547.6	-6547.6	-6448.1	-6440.4	-6433.5	-6426.6	-6423.6	-6423.1	-6423.0	-6420.7	-6184.2
F	5.9	7.8	7.6	6.6	6.0	5.8	5.6	5.2	5.1	5.0	5.6
Ind-Chi-Squared											1.5
Ind-P Value											0.158
Reg-Chi-Squared											4
Reg-P Value											0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$

Dependent variable is bribe amount in '000 Naira

Table A.28: Estimations Using Heckman Two-Step Procedure

	m1	m2	m3	m9 (OLS)
main				
tax_percentage	-0.031	-0.020	-1.045	0.051
	(0.124)	(0.124)	(1.507)	(0.032)
eprofit	0.170***	0.170***	0.153***	0.094***
	(0.008)	(0.008)	(0.040)	(0.025)
capital_labour	0.006***	0.007***	-0.004	0.006**
	(0.002)	(0.002)	(0.017)	(0.002)
un_sunk_cost	-165.540*	-166.170*	552.696	
	(91.874)	(92.014)	(1076.681)	
competitors_2	-0.823	-1.477	34.256	-2.717
	(13.796)	(13.830)	(79.221)	(7.437)
competitors_3	-5.508	-5.457	24.482	-1.674
	(8.666)	(8.687)	(57.723)	(4.518)
competitors_4	-11.208	-11.872	55.515	-7.244*
	(10.272)	(10.300)	(103.909)	(4.210)
gov_customer	20.929	21.050	-55.304	16.781
	(14.112)	(14.145)	(123.849)	(14.474)
regulation_realtime	-0.594	-0.569	-3.223	0.013
	(0.493)	(0.494)	(4.241)	(0.190)
eexternal_consultant_naira	0.605***	0.597***	1.273	0.027
	(0.153)	(0.153)	(1.122)	(0.184)
african	-3.637	-4.017	54.564	-6.632**
	(9.132)	(9.151)	(90.450)	(3.041)
trade		-24.885**		-16.972***
		(11.805)		(6.381)
infraserv			-30.058	1.160
			(43.219)	(2.314)
eun_sunk_cost				-88.986***
				(24.672)
Constant	-132.563	-134.154	689.522	5.397
	(97.258)	(97.407)	(1218.583)	(6.422)
bribe_dummy				
infraserv	0.174***	0.174***	0.174***	
	(0.058)	(0.058)	(0.058)	
trade	-0.025	-0.025	-0.025	
	(0.225)	(0.225)	(0.225)	
tax_percentage	0.006***	0.006***	0.006***	
	(0.001)	(0.001)	(0.001)	
eprofit	0.000	0.000	0.000	
	(0.000)	(0.000)	(0.000)	
capital_labour	0.000*	0.000*	0.000*	
	(0.000)	(0.000)	(0.000)	
un_sunk_cost	-4.260***	-4.260***	-4.260***	
	(1.127)	(1.127)	(1.127)	
competitors_2	-0.233	-0.233	-0.233	
	(0.280)	(0.280)	(0.280)	
competitors_3	-0.204	-0.204	-0.204	
	(0.177)	(0.177)	(0.177)	
competitors_4	-0.428**	-0.428**	-0.428**	
	(0.167)	(0.167)	(0.167)	
gov_customer	0.504*	0.504*	0.504*	
	(0.265)	(0.265)	(0.265)	
regulation_realtime	0.017**	0.017**	0.017**	
	(0.007)	(0.007)	(0.007)	
eexternal_consultant_naira	-0.004*	-0.004*	-0.004*	
	(0.002)	(0.002)	(0.002)	
african	-0.382***	-0.382***	-0.382***	
	(0.146)	(0.146)	(0.146)	
Constant	-3.530***	-3.530***	-3.530***	
	(1.027)	(1.027)	(1.027)	
Inverse-Mills Ratio	16	17	-253	
	(28.1)	(28.2)	(391.4)	
Observations	2001	2001	2001	2001
Censored Observations	972	972	972	
Chi-Squared	711.3	713.4	38.2	
F-Stat				4.83

$*p < 0.10, **p < 0.05, ***p < 0.01$

Dependent variable of selection equation is a dummy variable equal to 1 if the firm reported a positive bribe payment; 0 otherwise. Dependent variable of outcome equation is bribe amount in '000 Naira.

Table A.29: IV Regressions

	Model 1 (2SLS)	Model 2 (GMM)
eprofit	0.195***	0.000***
	(0.072)	(0.000)
employee_total	0.202	0.014***
	(0.156)	(0.002)
ecapital_stock	0.006	-0.000
	(0.004)	(0.000)
eun_sunk_cost	-154.942***	-4.419***
	(53.232)	(0.248)
infraserv	-3.449	0.043***
	(4.575)	(0.013)
trade	-25.196**	-0.102
	(10.454)	(0.080)
tax_percentage	-0.063	0.004***
	(0.053)	(0.000)
Constant	-2.858	
	(14.371)	
Observations	1029.000	1029.000
Chi-Squared	34.61099	

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dependent variable is bribe amount in '000 Naira. 2SLS: Identifying Instruments: Dummy=1 if sole owner/majority shareholder has university level education; Dummy=1 if firm is foreign owned; age of firm; and amount (in '000 Naira) spent on security. GMM: Identifying Instruments: Grouped (Industry-Location) Average Profit Per Employee ('000 Naira).

Table A.30: Bribery, by Region (NBS Data)

region	Dummy=1 if firm admitted to bribing; 0 otherwise No	Yes	Total	Sample Size
Abia Freq	5.0	3.0	8.0	8
Cell%	1.5	0.9	2.4	
Row%	62.5	37.5	100.0	
Col%	1.9	3.9	2.4	
Adamawa Freq	2.0	1.0	3.0	3
Cell%	0.6	0.3	0.9	
Row%	66.7	33.3	100.0	
Col%	0.8	1.3	0.9	
Akwa ibom Freq	3.0	0.0	3.0	3
Cell%	0.9	0.0	0.9	
Row%	100.0	0.0	100.0	
Col%	1.2	0.0	0.9	
Anambra Freq	6.0	3.0	9.0	9
Cell%	1.8	0.9	2.7	
Row%	66.7	33.3	100.0	
Col%	2.3	3.9	2.7	
Bauchi Freq	10.0	1.0	11.0	11
Cell%	3.0	0.3	3.3	
Row%	90.9	9.1	100.0	
Col%	3.8	1.3	3.3	
Bayelsa Freq	3.0	0.0	3.0	3
Cell%	0.9	0.0	0.9	
Row%	100.0	0.0	100.0	
Col%	1.2	0.0	0.9	
Benue Freq	2.0	0.0	2.0	2
Cell%	0.6	0.0	0.6	
Row%	100.0	0.0	100.0	
Col%	0.8	0.0	0.6	
Borno Freq	1.0	2.0	3.0	3
Cell%	0.3	0.6	0.9	
Row%	33.3	66.7	100.0	
Col%	0.4	2.6	0.9	
Cross river Freq	1.0	0.0	1.0	1
Cell%	0.3	0.0	0.3	
Row%	100.0	0.0	100.0	
Col%	0.4	0.0	0.3	
Delta Freq	4.0	4.0	8.0	8
Cell%	1.2	1.2	2.4	
Row%	50.0	50.0	100.0	
Col%	1.5	5.3	2.4	
Ebonyi Freq	1.0	2.0	3.0	3
Cell%	0.3	0.6	0.9	
Row%	33.3	66.7	100.0	
Col%	0.4	2.6	0.9	
Edo Freq	6.0	2.0	8.0	8
Cell%	1.8	0.6	2.4	
Row%	75.0	25.0	100.0	
Col%	2.3	2.6	2.4	
Ekiti Freq	7.0	1.0	8.0	8
Cell%	2.1	0.3	2.4	
Row%	87.5	12.5	100.0	
Col%	2.7	1.3	2.4	
Enugu Freq	3.0	1.0	4.0	4
Cell%	0.9	0.3	1.2	
Row%	75.0	25.0	100.0	
Col%	1.2	1.3	1.2	
Gombe Freq	3.0	1.0	4.0	4
Cell%	0.9	0.3	1.2	
Row%	75.0	25.0	100.0	
Col%	1.2	1.3	1.2	

Table A.31: Bribery, by Region (NBS Data) (cont'd)

region	Dummy=1 if firm admitted to bribing; 0 otherwise No	Yes	Total	Sample Size
Imo Freq	2.0	0.0	2.0	2
Cell%	0.6	0.0	0.6	
Row%	100.0	0.0	100.0	
Col%	0.8	0.0	0.6	
Jigawa Freq	8.0	0.0	8.0	8
Cell%	2.4	0.0	2.4	
Row%	100.0	0.0	100.0	
Col%	3.1	0.0	2.4	
Kaduna Freq	9.0	1.0	10.0	10
Cell%	2.7	0.3	3.0	
Row%	90.0	10.0	100.0	
Col%	3.5	1.3	3.0	
Kano Freq	19.0	3.0	22.0	22
Cell%	5.7	0.9	6.5	
Row%	86.4	13.6	100.0	
Col%	7.3	3.9	6.5	
Katsina Freq	10.0	4.0	14.0	14
Cell%	3.0	1.2	4.2	
Row%	71.4	28.6	100.0	
Col%	3.8	5.3	4.2	
Kebbi Freq	3.0	0.0	3.0	3
Cell%	0.9	0.0	0.9	
Row%	100.0	0.0	100.0	
Col%	1.2	0.0	0.9	
Kogi Freq	3.0	0.0	3.0	3
Cell%	0.9	0.0	0.9	
Row%	100.0	0.0	100.0	
Col%	1.2	0.0	0.9	
Kwara Freq	1.0	4.0	5.0	5
Cell%	0.3	1.2	1.5	
Row%	20.0	80.0	100.0	
Col%	0.4	5.3	1.5	
Lagos Freq	103.0	34.0	137.0	137
Cell%	30.7	10.1	40.8	
Row%	75.2	24.8	100.0	
Col%	39.6	44.7	40.8	
Nassarawa Freq	4.0	1.0	5.0	5
Cell%	1.2	0.3	1.5	
Row%	80.0	20.0	100.0	
Col%	1.5	1.3	1.5	
Niger Freq	1.0	2.0	3.0	3
Cell%	0.3	0.6	0.9	
Row%	33.3	66.7	100.0	
Col%	0.4	2.6	0.9	
Ogun Freq	8.0	1.0	9.0	9
Cell%	2.4	0.3	2.7	
Row%	88.9	11.1	100.0	
Col%	3.1	1.3	2.7	
Ondo Freq	3.0	0.0	3.0	3
Cell%	0.9	0.0	0.9	
Row%	100.0	0.0	100.0	
Col%	1.2	0.0	0.9	
Osun Freq	7.0	0.0	7.0	7
Cell%	2.1	0.0	2.1	
Row%	100.0	0.0	100.0	
Col%	2.7	0.0	2.1	
Oyo Freq	4.0	1.0	5.0	5
Cell%	1.2	0.3	1.5	
Row%	80.0	20.0	100.0	
Col%	1.5	1.3	1.5	

Table A.32: Bribery, by Region (NBS Data) (cont'd)

region	**No**	**Yes**	**Total**	**Sample Size**
	Dummy=1 if firm admitted to bribing; 0 otherwise			
Plateau Freq	6.0	2.0	8.0	8
Cell%	1.8	0.6	2.4	
Row%	75.0	25.0	100.0	
Col%	2.3	2.6	2.4	
Sokoto Freq	3.0	0.0	3.0	3
Cell%	0.9	0.0	0.9	
Row%	100.0	0.0	100.0	
Col%	1.2	0.0	0.9	
Taraba Freq	1.0	0.0	1.0	1
Cell%	0.3	0.0	0.3	
Row%	100.0	0.0	100.0	
Col%	0.4	0.0	0.3	
Yobe Freq	3.0	0.0	3.0	3
Cell%	0.9	0.0	0.9	
Row%	100.0	0.0	100.0	
Col%	1.2	0.0	0.9	
Zamfara Freq	2.0	1.0	3.0	3
Cell%	0.6	0.3	0.9	
Row%	66.7	33.3	100.0	
Col%	0.8	1.3	0.9	
FCT (Abuja) Freq	3.0	1.0	4.0	4
Cell%	0.9	0.3	1.2	
Row%	75.0	25.0	100.0	
Col%	1.2	1.3	1.2	
Total Freq	260.0	76.0	336.0	336
Cell%	77.4	22.6	100.0	
Row%	77.4	22.6	100.0	
Col%	100.0	100.0	100.0	

Table A.33: Summary Statistics Of Bribery And Meeting With Officials - NBS Data (Manu. and Full)

	(1)	(2)	(3)	(4)	(5)	(6)
	Manufacturing: Bribery/ Meeting With Official			Full Sample: Bribery/ Meeting With Official		
	All Firms	Met With An Official	Bribing Firms	All Firms	Met With An Official	Bribing Firms
Met With An Official	0.37	1	1	0.31	1	1
(Mean) Average Number Of Types Of Officials Met With	1.66 (2.72)	4.43 (2.74)	4.20 (2.50)	1.09 (2.07)	3.49 (2.32)	3.49 (2.33)
Bribery Episode	0.14	0.36	1	0.10	0.33	1
(Mean)Number Of Types Of Bribery Episodes	0.44 (1.39)	1.16 (2.07)	3.20 (2.31)	0.28 (1.07)	0.90 (1.76)	2.67 (2.12)
Observations	331	124	45	2089	651	218
(Mean)Value Of Bribes ('000 ₦)(₦)	0.616 (4.56)	1.65 (7.36)	3.43 (9.49)	0.71 (4.55)	2.28 (7.93)	6.29 (12.25)

The unit of observation is the firm. Standard deviations are in parentheses.

Table A.34: Cross-Tabulations Of The Payment Of Bribes, By Interactions With Public Officials (NBS Data)

Conda: Clearing goods through customs

Operation Performed \ Bribe Paid	No	Yes	Total
No	277	0	277
Yes	41	18	59
Total	318	18	336

P(bribea|conda)=(18/336)÷(59/336)=0.31

Condb: Obtaining Road Worthy Certificates

Operation Performed \ Bribe Paid	No	Yes	Total
No	260	0	260
Yes	61	15	76
Total	321	15	336

P(bribeb|condb)=(15/336)÷(76/336)=0.20

Condc: Procurement of goods and services from the government

Operation Performed \ Bribe Paid	No	Yes	Total
No	310	0	310
Yes	20	6	26
Total	330	6	336

P(bribec|condc)=(6/336)÷(26/336)=0.26

Condd: Obtaining business licenses and permits

Operation Performed \ Bribe Paid	No	Yes	Total
No	274	0	274
Yes	46	16	62
Total	320	16	336

P(bribed|condd)=(16/336)÷(62/336)=0.3

Conde: Procurement of goods and services from private companies

Operation Performed \ Bribe Paid	No	Yes	Total
No	307	0	307
Yes	26	3	29
Total	333	3	336

P(bribee|conde)=(3/336)÷(29/336)=0.10

Condf: Getting clearance for environmental or sanitary regulations

Operation Performed \ Bribe Paid	No	Yes	Total
No	273	0	273
Yes	44	19	63
Total	317	19	336

P(bribef|condf)=(19/336)÷(63/336)=0.30

Condg: Residence and work permits

Operation Performed \ Bribe Paid	No	Yes	Total
No	285	0	285
Yes	42	9	51
Total	327	9	336

P(bribeg|condg)=(9/336)÷(51/336)=0.18

Table A.35: Cross-Tabulations Of The Payment Of Bribes, By Interactions With Public Officials (NBS Data) (cont'd)

Condh: Vehicle regulations

Operation Performed \ Bribe Paid	No	Yes	Total
No	259	0	259
Yes	57	20	77
Total	316	20	336

P(bribeh|condh)=(20/336)÷(77/336)=0.26

Condi: Police Investigations

Operation Performed \ Bribe Paid	No	Yes	Total
No	295	0	295
Yes	22	19	41
Total	317	19	336

P(bribei|condi)=(19/336)÷(41/336)=0.46

Condj: Traffic Offences

Operation Performed \ Bribe Paid	No	Yes	Total
No	275	0	275
Yes	30	31	61
Total	305	31	336

P(bribej|condj)=(31/336)÷(61/336)=0.51

Condk: Contact with the court

Operation Performed \ Bribe Paid	No	Yes	Total
No	312	0	312
Yes	17	7	24
Total	329	7	336

P(bribek|condk)=(7/336)÷(24/336)=0.29

Table A.36: Probit Estimations On The Propensity To Bribe - NBS Data

	0	1	2	3	4	5	6	7	8	9	10	11	12
employee	-0.031	-0.031	-0.051	-0.034	-0.038	-0.029	-0.032	-0.032	-0.045	-0.053	-0.091	-0.029	-0.074
	(0.090)	(0.090)	(0.093)	(0.091)	(0.092)	(0.090)	(0.092)	(0.090)	(0.095)	(0.090)	(0.100)	(0.090)	(0.102)
trade	0.907***		0.699**	0.763***	0.764***	0.889***	0.577**	0.867***	0.501*	0.624**	0.529*	0.929***	0.549
	(0.223)		(0.289)	(0.247)	(0.254)	(0.239)	(0.279)	(0.269)	(0.280)	(0.253)	(0.278)	(0.238)	(0.352)
foreign	-0.693***	-0.693***	-0.659***	-0.669***	-0.714***	-0.687***	-0.696***	-0.700***	-0.686***	-0.721***	-1.031***	-0.692***	-1.074***
	(0.229)	(0.229)	(0.234)	(0.233)	(0.232)	(0.231)	(0.235)	(0.232)	(0.241)	(0.236)	(0.289)	(0.230)	(0.311)
lage	-0.073	-0.073	-0.072	-0.061	-0.065	-0.078	-0.071	-0.072	-0.075	-0.058	-0.043	-0.076	-0.073
	(0.099)	(0.099)	(0.099)	(0.097)	(0.099)	(0.100)	(0.101)	(0.099)	(0.100)	(0.100)	(0.108)	(0.099)	(0.116)
conda		0.907***											
		(0.223)											
condb			0.391										-0.844*
			(0.256)										(0.475)
condc				0.466									-0.216
				(0.327)									(0.426)
condd					0.405*								0.171
					(0.245)								(0.382)
conde						0.076							-0.109
						(0.322)							(0.450)
condf							0.614**						0.355
							(0.261)						(0.369)
condg								0.090					-0.789**
								(0.286)					(0.357)
condh									0.789***				1.067***
									(0.240)				(0.375)
condi										0.726**			0.500
										(0.283)			(0.389)
condj											1.678***		2.048***
											(0.256)		(0.348)
condk												-0.146	-1.067**
												(0.364)	(0.534)
Constant	-0.751	-0.751	-0.841	-0.775	-0.879	-0.744	-0.941	-0.753	-0.967*	-0.751	-1.261**	-0.748	-1.627***
	(0.654)	(0.654)	(0.613)	(0.655)	(0.610)	(0.655)	(0.586)	(0.655)	(0.573)	(0.658)	(0.508)	(0.654)	(0.462)
State Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Pseudo R-squared	0.150	0.150	0.159	0.157	0.159	0.150	0.168	0.150	0.188	0.170	0.310	0.150	0.383
Observations	287.000	287.000	287.000	287.000	287.000	287.000	287.000	287.000	287.000	287.000	287.000	287.000	287.000
Log-Likelihood	-137	-137	-135	-135	-135	-136	-134	-136	-130	-133	-111	-136	-99
Chi-Squared	50.1	50.1	57.5	51.3	53.3	51.5	57.1	50.2	66.2	51.5	73.2	50.0	104.6
Reg-Chi-Squared	21.78	21.78	21.94	23.15	21.15	21.72	20.96	21.53	22.18	20.25	30.94	21.87	39.12
Reg-P Value	0.3	0.3	0.3	0.2	0.3	0.3	0.3	0.3	0.3	0.4	0.0	0.3	0.0

$*p < 0.10, **p < 0.05, ***p < 0.01$

Dependent variable is a dummy variable equal to 1 if the firm reported paying a bribe; 0 otherwise.

Table A.37: Censored Probit Models - NBS Data

Dependent Variable	Conda	Bribea	Condb	Bribeb	Condc	Bribec	Condd	Bribed	Conde	Bribee
employee	0.136**	0.969	-0.048	0.244	0.017	0.116	-0.028	-0.180	-0.125**	1.619
	(0.057)	(1.779)	(0.048)	(0.177)	(0.062)	(0.260)	(0.052)	(0.171)	(0.063)	(1.710)
foreign	0.416***	2.430	-0.190	0.419	-0.184	1.986*	-0.105	0.091	-0.130	2.477
	(0.144)	(5.451)	(0.139)	(0.596)	(0.170)	(1.146)	(0.138)	(0.405)	(0.167)	(1.607)
lage	0.017	-0.107	0.075	0.069	0.072	-0.316	-0.053	-0.115	0.105*	-1.539
	(0.059)	(0.295)	(0.051)	(0.273)	(0.070)	(0.621)	(0.052)	(0.207)	(0.061)	(1.677)
trade			1.638***	-3.651	1.629***	13.030**	1.434***	1.844	1.513***	-23.999
			(0.137)	(4.653)	(0.161)	(5.324)	(0.140)	(2.373)	(0.160)	(20.649)
Constant	-6.704***	-14.147	-6.255***	3.302	-0.959*	-24.024	-5.997*	1.352	-0.915*	26.230
	(1.240)	(29.120)	(1.972)	(5.867)	(0.492)	(.)	(3.429)	(2.287)	(0.474)	(33.770)
Inverse-Mills Ratio		8.806		-2.968		10.520**		2.125		-22.210
		(17.385)		(4.185)		(4.522)		(2.465)		(18.787)
Industry Dummies	YES									
State Dummies	YES									
Region-State Dummies	YES									
Pseudo R-squared	0.109	0.115	0.193	0.176	0.241	0.410	0.177	0.083	0.195	0.178
Observations	864	88	1210	119	904	47	1198	109	997	25
Log-Likelihood	-362	-49	-515	-61	-296	-19	-529	-65	-337	-13

$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A.38: Censored Probit Models - NBS Data (cont'd)

Dependent Variable	Condf	Bribef	Condg	Bribeg	Condh	Bribeh	Condi	Bribei	Condj	Bribej	Condk	Bribek
employee	0.045	-0.057	-0.000	-0.109	-0.011	-0.183	0.148***	-0.106	-0.004	-0.008	0.199***	-0.628
	(0.055)	(0.191)	(0.056)	(0.200)	(0.047)	(0.144)	(0.052)	(0.235)	(0.050)	(0.174)	(0.066)	(0.468)
foreign	-0.263*	0.316	0.171	-1.032	-0.172	0.320	-0.141	-0.266	0.037	-0.125	-0.166	0.546
	(0.144)	(0.511)	(0.145)	(0.767)	(0.126)	(0.533)	(0.149)	(0.492)	(0.136)	(0.418)	(0.181)	(0.853)
lage	0.039	-0.360	-0.078	0.337	0.089*	0.249	-0.003	-0.575**	0.121**	0.239	0.161**	-1.447
	(0.053)	(0.233)	(0.060)	(0.250)	(0.050)	(0.274)	(0.056)	(0.234)	(0.058)	(0.768)	(0.077)	(0.952)
trade	1.604***	-1.619	1.514***	-12.085***	1.649***	1.438	1.613***	-2.464	1.164***	-0.821	1.423***	-4.947**
	(0.151)	(2.067)	(0.140)	(3.706)	(0.137)	(4.023)	(0.147)	(1.658)	(0.132)	(6.777)	(0.187)	(2.084)
Constant	-0.929**	9.105***	-6.090***	12.767***	-0.988**	-8.228	-0.959**	11.392**	-1.096**	-0.919	-6.753***	13.371**
	(0.462)	(2.997)	(0.999)	(4.417)	(0.466)	(.)	(0.464)	(5.403)	(0.471)	(12.667)	(2.265)	(6.036)
Inverse Mills Ratio		-1.603		-10.390***		1.310		-2.831*		0.388		-4.942**
		(1.982)		(3.496)		(3.634)		(1.484)		(7.936)		(2.480)
Industry Dummies	YES											
State Dummies	YES											
Region-State Dummies	YES											
Pseudo R-squared	0.202	0.160	0.219	0.268	0.179	0.128	0.214	0.149	0.143	0.214	0.256	0.315
Observations	1047	98	945	81	1337	122	1083	82	1087	91	853	20
Log-Likelihood	-447	-55	-354	-36	-610	-68	-421	-48	-465	-49	-269	-9

$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A.39: Censored Probit Models - 1st Stage

	Model 1a	Model 2a	Model 3a	Model 4a	Model 5a
industry1	-0.248	0.703***	0.118	-0.145	0.048
	(0.355)	(0.246)	(0.251)	(0.225)	(0.270)
industry2	0.666			0.082	
	(0.594)			(0.573)	
industry3	0.155	0.252	-0.105	0.248	0.318
	(0.390)	(0.348)	(0.370)	(0.268)	(0.322)
industry4	0.941***	1.036***	0.212	0.381**	0.346
	(0.232)	(0.215)	(0.206)	(0.169)	(0.213)
industry5	-0.261	0.850***	0.218	-0.106	0.175
	(0.452)	(0.274)	(0.292)	(0.274)	(0.311)
industry6	0.410	0.871***	0.346	0.425**	0.056
	(0.288)	(0.249)	(0.248)	(0.207)	(0.284)
industry7	0.594**	0.629***	0.197	0.394**	0.450**
	(0.240)	(0.223)	(0.212)	(0.173)	(0.215)
industry8	0.285	0.513**	-0.135	0.293	0.228
	(0.283)	(0.251)	(0.271)	(0.200)	(0.253)
industry9	0.570**	0.631***	-0.047	0.188	0.245
	(0.245)	(0.227)	(0.228)	(0.181)	(0.225)
industry10	0.202	0.374	-0.130	0.086	0.140
	(0.279)	(0.250)	(0.258)	(0.200)	(0.249)
industry11	0.430	0.555*	0.396	0.164	0.366
	(0.333)	(0.302)	(0.288)	(0.262)	(0.303)
industry12	0.293	0.403	0.407	-0.080	0.108
	(0.324)	(0.292)	(0.266)	(0.259)	(0.306)
industry13	0.211	0.428*	0.349	-0.372	0.091
	(0.280)	(0.248)	(0.228)	(0.231)	(0.254)
industry14	-0.469	0.613***	-0.105	0.106	0.108
	(0.340)	(0.229)	(0.236)	(0.186)	(0.235)
Constant	-1.886***	-1.786***	-1.627***	-1.303***	-1.701***
	(0.217)	(0.201)	(0.180)	(0.149)	(0.189)
Pseudo R-squared	0.075	0.033	0.017	0.023	0.011
Observations	2089	2080	2080	2089	2080
Log-Likelihood	-508	-773	-502	-797	-525
Chi-Squared	69.5	49.3	17.2	34.3	11.6
P	0.000	0.000	0.190	0.002	0.559
Industry-Chi-Squared	63.7	.	.	32.3	.
Industry-P Value	0.000	.	.	0.002	.

$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A.40: Censored Probit Models - 2nd Stage

	Model 1b	Model 2b	Model 3b	Model 4b	Model 5b
foreign	-0.635	5.539	-6.814	-2.307**	8.031
	(1.244)	(.)	(.)	(1.153)	(.)
age	-0.002	0.272***	-0.081***	-0.008	-1.107***
	(0.018)	(0.017)	(0.016)	(0.022)	(0.008)
sum_bribe	2.885***	15.997	19.625***	2.137***	23.379
	(0.658)	(.)	(0.060)	(0.394)	(.)
sum_cond	-1.966***	-3.342***	-4.253***	-1.454***	-0.482***
	(0.423)	(0.246)	(0.058)	(0.324)	(0.047)
employee	0.031	-5.994	1.191	-0.250	4.687
	(0.268)	(.)	(.)	(0.431)	(.)
male	-2.159**	-0.279	-14.645	0.878	-8.864
	(0.929)	(.)	(.)	(0.636)	(.)
lambdaa	5.103***				
	(1.145)				
trade		-3.312	2.892	1.356*	-44.995
		(.)	(.)	(0.741)	(.)
lambdab		10.607***			
		(0.398)			
lambdac			110.897***		
			(0.166)		
lambdad				-1.217	
				(2.148)	
lambdae					-61.679
					(.)
Region Dummies			YES		
Constant	-10.296***	-19.315	-211.393	1.495	93.449
	(1.921)	(.)	(.)	(3.235)	(.)
Pseudo R-squared	0.869	1.000	1.000	0.862	1.000
Observations	118	224	99	205	95
Log-Likelihood	-10	-0	-0	-17	-0
Chi-Squared	.	.	.	.	.
P	.	.	.	.	.
Industry-Chi-Squared					
Industry-P Value					

$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A.41: Censored Probit Models - 1st Stage (Cont'd)

	Model 6a	Model 7a	Model 8a	Model 9a	Model 10a	Model 11a
industry1	-0.114	-0.181	0.301	-0.392	-0.131	-0.395
	(0.235)	(0.316)	(0.206)	(0.241)	(0.244)	(0.273)
industry3	0.012	0.732**	0.197	-0.273	-0.286	0.063
	(0.304)	(0.300)	(0.277)	(0.319)	(0.362)	(0.306)
industry4	0.485***	0.754***	0.597***	0.059	0.512***	-0.057
	(0.176)	(0.218)	(0.170)	(0.171)	(0.180)	(0.193)
industry5	0.086	0.118	0.528**	0.325	0.510**	0.305
	(0.269)	(0.337)	(0.235)	(0.236)	(0.246)	(0.258)
industry6	0.169	0.445*	0.353*	-0.080	0.220	
	(0.228)	(0.267)	(0.214)	(0.229)	(0.231)	
industry7	0.323*	0.428*	0.275	-0.144	0.041	-0.396*
	(0.182)	(0.228)	(0.179)	(0.183)	(0.195)	(0.218)
industry8	0.479**	0.420	0.338*	-0.213	-0.316	-0.708**
	(0.203)	(0.256)	(0.203)	(0.222)	(0.259)	(0.322)
industry9	-0.002	0.495**	0.462***	0.241	0.479**	0.049
	(0.196)	(0.230)	(0.179)	(0.176)	(0.188)	(0.200)
industry10	0.067	0.278	0.259	-0.067	-0.062	0.157
	(0.211)	(0.255)	(0.198)	(0.203)	(0.225)	(0.213)
industry11	-0.061	-0.307	0.368	0.029	0.215	0.390
	(0.298)	(0.451)	(0.253)	(0.268)	(0.277)	(0.263)
industry12	-0.519	0.489*	0.323	0.293	0.149	0.225
	(0.341)	(0.286)	(0.236)	(0.228)	(0.259)	(0.253)
industry13	-0.156	0.235	0.272	0.056	-0.301	-0.002
	(0.226)	(0.259)	(0.199)	(0.199)	(0.246)	(0.222)
industry14	0.245	0.104	0.219	-0.219	-0.069	-0.547**
	(0.191)	(0.249)	(0.187)	(0.195)	(0.208)	(0.246)
industry2					0.225	
					(0.576)	
Constant	-1.395***	-1.786***	-1.348***	-1.261***	-1.446***	-1.446***
	(0.156)	(0.201)	(0.152)	(0.146)	(0.161)	(0.161)
Pseudo R-squared	0.029	0.037	0.012	0.018	0.043	0.040
Observations	2080	2080	2080	2080	2089	1980
Log-Likelihood	-717	-572	-904	-673	-650	-448
Chi-Squared	39.9	39.9	21.7	24.0	55.7	33.5
P	0.000	0.000	0.061	0.031	0.000	0.001
Industry-Chi-Squared	.	.	.	.	54.1	.
Industry-P Value	.	.	.	.	0.000	.

$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A.42: Censored Probit Models - 2nd Stage (Cont'd)

	Model 6b	Model 7b	Model 8b	Model 9b	Model 10b	Model 11b
employee	1.109*	-9.681***	-0.636**	0.045	0.544***	-3.409
	(0.585)	(2.536)	(0.287)	(0.152)	(0.211)	(.)
trade	-4.546***	-13.172***	-0.099	0.299	-2.255**	-10.086
	(1.378)	(2.542)	(0.782)	(0.494)	(1.028)	(.)
foreign	-0.420	7.934*	0.417	-0.650*	-1.219**	3.902
	(.)	(4.441)	(0.610)	(0.380)	(0.595)	(.)
age	-0.078	0.052**	0.017	0.012	0.013	0.003
	(0.052)	(0.024)	(0.012)	(0.008)	(0.017)	(0.008)
sum_bribe	6.574***	16.591***	1.776***	1.011***	1.559***	10.262
	(2.121)	(2.659)	(0.271)	(0.177)	(0.280)	(.)
sum_cond	-1.444**	-9.626***	-0.705***	-0.447***	-1.057***	-7.860***
	(0.621)	(1.460)	(0.159)	(0.149)	(0.170)	(0.077)
male	1.898	16.882***	-2.031***	-0.518	0.456	-2.985
	(1.305)	(3.754)	(0.704)	(0.471)	(0.828)	(.)
lambdaf	-1.873					
	(1.800)					
lambdag		35.820***				
		(6.636)				
lambdah			-4.085**			
			(1.728)			
lambdai				1.409*		
				(0.738)		
lambdaj					-1.928**	
					(0.905)	
lambdak						18.172
						(.)
Constant	-0.078	-70.104***	5.202*	-1.738	1.706	-32.454
	(3.063)	(12.615)	(2.791)	(1.732)	(1.752)	(.)
Region Dummies	YES					
Pseudo R-squared	0.953	0.930	0.856	0.719	0.796	1.000
Observations	193	133	271	184	179	74
Log-Likelihood	-6	-5	-21	-35	-25	-0
Chi-Squared	.	.	77.0	77.7	61.6	.
P	.	.	0.000	0.000	0.000	.
Industry-Chi-Squared						
Industry-P Value						

$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A.43: Multinomial Logit Models On Perception Of Corruption As An Obstacle To Business Operations - ES Data

	(Model 1 - log-odds ratios)					(Model 1 - odds ratios)				
	no obstacle (base category)	minor obstacle	moderate obstacle	major obstacle	very severe obstacle	no obstacle (base category)	minor obstacle	moderate obstacle	major obstacle	very severe obstacle
infraserv	-	-0.209	-0.109	0.030	-0.045	-	-0.209	-0.109	0.030	-0.045
	-	(0.143)	(0.139)	(0.141)	(0.191)	-	(0.143)	(0.139)	(0.141)	(0.191)
trade	-	0.745	0.352	0.450	0.730	-	0.745	0.352	0.450	0.730
	-	(0.563)	(0.555)	(0.593)	(0.651)	-	(0.563)	(0.555)	(0.593)	(0.651)
tax percentage	-	0.002	0.014***	0.006**	0.002	-	0.002	0.014***	0.006**	0.002
	-	(0.002)	(0.002)	(0.003)	(0.003)	-	(0.002)	(0.002)	(0.003)	(0.003)
profit	-	0.000	0.000*	0.000	0.000	-	0.000	0.000*	0.000	0.000
	-	(0.000)	(0.000)	(0.000)	(0.000)	-	(0.000)	(0.000)	(0.000)	(0.000)
capital labour	-	0.000	-0.000	0.000	0.000	-	0.000	-0.000	0.000	0.000
	-	(0.000)	(0.000)	(0.000)	(0.000)	-	(0.000)	(0.000)	(0.000)	(0.000)
un_sunk_cost	-	1.865	-4.359*	-0.422	-1.198	-	1.865	-4.359*	-0.422	-1.198
	-	(2.642)	(2.573)	(2.672)	(3.222)	-	(2.642)	(2.573)	(2.672)	(3.222)
competitors (2)	-	0.387	-0.014	-0.477	-1.959**	-	0.387	-0.014	-0.477	-1.959**
	-	(0.651)	(0.715)	(0.710)	(0.895)	-	(0.651)	(0.715)	(0.710)	(0.895)
competitors (3)	-	-0.135	0.049	-0.282	-1.544***	-	-0.135	0.049	-0.282	-1.544***
	-	(0.465)	(0.472)	(0.444)	(0.425)	-	(0.465)	(0.472)	(0.444)	(0.425)
competitors (4)	-	-0.302	0.074	-0.445	-1.442***	-	-0.302	0.074	-0.445	-1.442***
	-	(0.446)	(0.455)	(0.426)	(0.392)	-	(0.446)	(0.455)	(0.426)	(0.392)
gov_customer	-	-1.067	-0.421	0.217	-0.268	-	-1.067	-0.421	0.217	-0.268
	-	(0.795)	(0.601)	(0.520)	(0.660)	-	(0.795)	(0.601)	(0.520)	(0.660)
regulation_realtime	-	0.033	0.038	0.061**	0.094***	-	0.033	0.038	0.061**	0.094***
	-	(0.025)	(0.025)	(0.024)	(0.024)	-	(0.025)	(0.025)	(0.024)	(0.024)
external_consultant_naira	-	-0.000	-0.001**	-0.000	-0.001**	-	-0.000	-0.001**	-0.000	-0.001**
	-	(0.000)	(0.000)	(0.000)	(0.000)	-	(0.000)	(0.000)	(0.000)	(0.000)
african	-	0.346	0.425	0.236	0.835*	-	0.346	0.425	0.236	0.835*
	-	(0.332)	(0.312)	(0.322)	(0.428)	-	(0.332)	(0.312)	(0.322)	(0.428)
Constant	-	0.897	-5.516**	-1.207	-1.222	-	0.897	-5.516**	-1.207	-1.222
	-	(2.396)	(2.401)	(2.431)	(2.927)	-	(2.396)	(2.401)	(2.431)	(2.927)
Industry Dummies			YES					YES		
Region Dummies			NO					NO		
Pseudo R-squared			0.029					0.029		
Observations			2001					2001		
Log-Pseudolikelihood			-3038.9					-3038.9		

$*p < 0.10, **p < 0.05, ***p < 0.01$

A.1.3 Estimations For The Tobit Specification Test

The following 2 tables show the estimations performed in conducting the specification test for the Tobit model.

Table A.44: Probit Estimations On All Firms

	p0	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10
Pseudo R-squared	0.014	0.014	0.014	0.016	0.017	0.023	0.025	0.027	0.028	0.031	0.199
Observations	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001
Log-Likelihood	-1366.4	-1366.7	-1366.7	-1363.9	-1362.2	-1354.0	-1351.4	-1349.4	-1347.4	-1343.4	-1110.8
L	-7092.4	-7092.8	-6834.1	-6825.4	-6817.3	-6808.2	-6804.4	-6801.9	-6788.1	-6784	-6468.5
Lrt	-1089.6	-1090.4	-772	-770	-767.6	-763.2	-761.6	-757.6	-730.2	-726.6	-568.6

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dependent Variable is Dummy=1 if bribe paid, 0 otherwise. L is the sum of the log-likelihood for the truncated tobit model and the log-likelihood for the probit model. This is the log-likelihood for the unrestricted model in the tobit specification test. LRT is the likelihood ratio test statistic for the tobit model specification test. Control variables are the same as those used in the truncated models in Table A.45

Table A.45: Truncated Tobit Estimations For Bribing Firms Only

	tt0	tt1	tt2	tt3	tt4	tt5	tt6	tt7	tt8	tt9	tt10
model											
infraserv	2.228	2.161	-4.309	-4.301	-2.739	-3.121	-2.992	-2.734	-2.491	-2.486	2.589
	(3.810)	(3.807)	(3.642)	(3.621)	(3.543)	(3.566)	(3.553)	(3.547)	(3.453)	(3.451)	(3.682)
import	-18.347										
	(12.432)										
export	-0.171										
	(22.321)										
tax_percentage	-0.086	-0.087	-0.072	-0.077	-0.067	-0.058	-0.054	-0.057	-0.064	-0.064	-0.122*
	(0.068)	(0.068)	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)	(0.063)
trade		-14.815	-21.103**	-25.617***	-23.351**	-24.103**	-23.773**	-23.685**	-22.821**	-22.811**	-12.515
		(11.350)	(9.578)	(9.466)	(9.696)	(9.624)	(9.624)	(9.574)	(8.892)	(8.855)	(7.747)
eprofit			0.180***	0.177***	0.175***	0.175***	0.175***	0.175***	0.167***	0.167***	0.155***
			(0.040)	(0.040)	(0.039)	(0.039)	(0.039)	(0.039)	(0.034)	(0.034)	(0.034)
capital_labour				0.006	0.006	0.006*	0.006*	0.006*	0.006*	0.006*	0.007**
				(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
eun_sunk_cost					-119.332***	-116.665***	-118.703***	-118.354***	-126.126***	-126.641***	
					(34.033)	(34.604)	(34.610)	(34.538)	(36.019)	(36.110)	
competitors_2						4.951	4.969	5.512	4.634	4.704	-0.217
						(8.796)	(8.789)	(8.775)	(8.743)	(8.738)	(9.617)
competitors_3						1.063	0.908	0.687	-0.775	-0.506	5.517
						(5.010)	(5.011)	(5.029)	(5.136)	(5.126)	(5.931)
competitors_4						-3.404	-3.775	-3.934	-4.838	-4.571	6.999
						(4.494)	(4.489)	(4.515)	(4.582)	(4.604)	(6.541)
gov_customer							17.946	18.120	19.400	19.392	15.720
							(15.855)	(15.913)	(16.193)	(16.209)	(14.424)
regulation_realtime								-0.380	-0.757**	-0.759**	0.081
								(0.266)	(0.380)	(0.380)	(0.456)
eexternal_consultant_naira									0.649	0.649	0.720
									(0.464)	(0.464)	(0.475)
african										-1.870	4.375
										(4.096)	(3.978)
un_sunk_cost											20.741
											(60.197)
Industry Dummies											YES
Region Dummies											YES
Constant	45.806***	45.892***	17.851***	17.124***	6.319	8.169	7.708	9.037	10.664	12.126	63.599
	(5.598)	(5.602)	(6.137)	(6.072)	(7.149)	(8.388)	(8.382)	(8.349)	(7.677)	(7.877)	(49.687)
sigma	63.161***	63.170***	49.125***	48.843***	48.543***	48.498***	48.442***	48.417***	47.865***	47.863***	44.160***
	(10.604)	(10.603)	(5.383)	(5.372)	(5.306)	(5.324)	(5.332)	(5.331)	(4.861)	(4.859)	(4.985)
Pseudo R-squared	0.000	0.000	0.045	0.046	0.048	0.048	0.048	0.048	0.050	0.050	0.065
Observations	1029	1029	1029	1029	1029	1029	1029	1029	1029	1029	1029
Log-Likelihood	-5726.0	-5726.1	-5467.4	-5461.5	-5455.1	-5454.2	-5453.0	-5452.5	-5440.7	-5440.6	-5357.7
F	1.0	1.1	6.2	5.6	5.1	5.0	4.7	4.5	4.7	4.4	14.3

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dependent Variable is Bribe ('000 Naira) per Employee.

A.1.4 The Independence Of Irrelevant Alternatives Assumption

Hausman & McFadden [1984] introduced a test of the IIA property of the MNL model. This test is related to the matrix version of the Hausman test [Hausman , 1978]. The Hausman-McFadden test obtains the parameter vector of the original MNL model, θ_j (for each j category), and a variance-covariance matrix, $V(\theta_j)$. The sample is then truncated by excluding one of the choices from the dependent variable (e.g. Sales %). A new parameter vector, $\hat{\theta}_j$, and variance-covariance matrix, $V(\hat{\theta}_j)$, are obtained from the truncated sample and the relevant test statistic is calculated as:

$$\tau = [\hat{\theta}_j - \theta_j]'[V(\hat{\theta}_j) - V(\theta_j)]^{-1}[\hat{\theta}_j - \theta_j] \sim \chi^2_k \qquad \text{(A.1.4.1)}$$

The null hypothesis is that the IIA condition holds; in other words, there is no difference in the estimate of the parameters if an arbitrary category is excluded. If the random exclusion of a category changes the parameter estimates, then this implies that the relative risk ratios are affected and are thus not independent. Both a necessary and sufficient condition for the test to generate an interpretable result is that the matrix derived from differencing the two variance-covariance matrices is positive definite.

If $\hat{\theta}_j$ were a fully efficient estimator, then computation of $[V(\hat{\theta}_j) - V(\theta_j)]^{-1}$ would be simple because in such case: $[V(\hat{\theta}_j) - V(\theta_j)]^{-1} = V(\theta_j) - V(\hat{\theta}_j)$. However, this would be inconsistent with the use of Huber/White robust standard errors in the estimation of the model since the use of such assumes that the variance-covariance matrix is not efficiently estimated. Using robust standard errors does not allow for easy computation of the variance-covariance matrix; a bootstrapped version of the Hausman test can be used instead [Cameron & Trivedi , 2005] without the efficiency requirement imposed by the standard Hausman test.

The differenced variance-covariance matrices, $V(\hat{\theta}_j) - V(\theta_j)$, is estimated using:

$$\hat{V}_{boot} = \frac{1}{B-1}\sum_{b=1}^{B}[\hat{\theta}_j - \theta_j - \overline{\theta_{diff}}][\hat{\theta}_j - \theta_j - \overline{\theta_{diff}}]' \qquad \text{(A.1.4.2)}$$

where $\overline{\theta_{diff}} = \frac{1}{B}\sum_b \hat{\theta}_j - \theta_j$.

To deal with the problem of a non-positive definite differenced matrix in the Hausman-McFadden test, the Small-Hsiao test [Small & Hsiao , 1985] is also used to test the IIA condition of the MNL model. This test is performed by running the initial model and taking the log-likelihood of the estimation, L_{full}. One category is then randomly removed and the estimation is run on the truncated sample, producing a second log-likelihood, L_{trunc}. This sample is then duplicated and added to the removed sample and the estimation is run again, producing a third log-likelihood value, L_{dup}. The test statistic is calculated as: $SH = -2[L_{dup} - (L_{full} + L_{trunc})] \sim \chi^2_k$. In order to improve the strength of the test results and to deal with the possibility of different excluded categories giving different test results, both tests are performed multiple times, each time removing a different category.

Results The differenced variance-covariance matrix of the adjusted Hausman test is not positive-definite, so this test could not be applied in the present case. Results from the Small-Hsiao test are shown in Table A.46. All results point to a rejection of the IIA assumption of the multinomial logit model, this result applies for both model 11 and 12 when performing the test by omitting each category respectively.

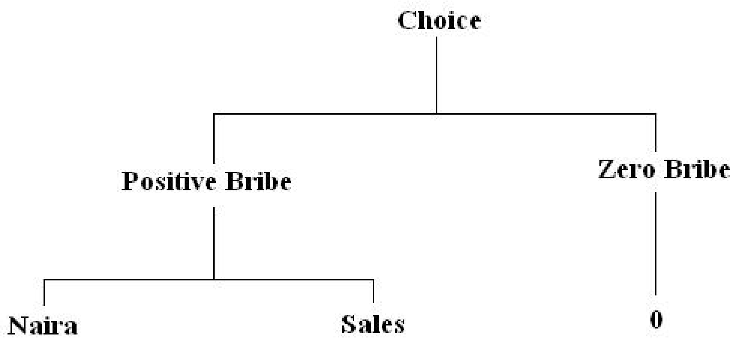
Table A.46: Small-Hsiao Tests For Independence Of Irrelevant Alternatives Assumption In Multinomial Logit Model

		Model 12 (k=56)
Full Model		
n=2001	L_{full}	-1493.3731
Omit Naira		
n=1774	L_{trunc}	-970.71938
n=3775	L_{dup}	-2594.9804
Small-Hsiao Test Statistic		261.77584
Result		Reject IIA
Omit Sales		
n=1199	L_{trunc}	-360.52027
n=3200	L_{dup}	-2242.1706
Small-Hsiao Test Statistic		776.55446
Result		Reject IIA
Omit Non-Bribers		
n=1029	L_{trunc}	-380.73979
n=3030	L_{dup}	-2317.6554
Small-Hsiao Test Statistic		887.08502
Result		Reject IIA

Results of the Small-Hsiao test do not change when using more sparse multinomial logit models.

Due to the rejection of the IIA assumption, it appears that another model should be used in order to model the responses to the question on bribery. The nested logit (NL) model is an alternative model that relaxes the IIA assumption. The NL model groups the alternatives into subgroups, allowing the variance of the errors to differ across sub-groups whilst maintaining the IIA assumption within each group. This might be more suited to the current dataset because of the nested nature of the data. For example, the firm must first decide whether or not to report a bribe; if it does report a bribe, then it must decide whether or not to report in terms of sales or Naira. This nested nature of the reporting decision is shown in Figure 17.

Figure 17: The Nested Nature Of The Bribe Decision



Denoting the alternatives by subscripts (j, k), where j represents the group (level 1) and k represents the choice (level 2) within the group. Figure 17 displays the 3 alternative choices responses within the 2 subgroups. One of the branches has only one choice associated with it, so the conditional probability, $P_{j|Zerobribe} = P_{0|Zerobribe} = 1$. The log-likelihood associated with the full information maximum likelihood estimation of the model is:

$$lnL = \sum_{i=1}^{n} = ln\left[P(choice|group)_i * P(group)_i\right] \tag{A.1.4.3}$$

A.2 Appendices For Chapter 2

A.2.1 Appendix A

The Unrelated Question Randomised Response Design The unrelated question RR design was introduced via papers by Horvitz, Shah, and Simmons [1967] and Greenberg, Abul-Ela, Simmons & Horvitz [1969]. The design of the procedure is similar to the one for the original design [Warner , 1965] described in Section 4.2 with the exception that the second question posed to the manager is entirely unrelated to the first (sensitive) question and is also unsensitive. For example:

Question 1: "Have you paid a bribe in the last 12 months?"

Question 2: "Have you used a bus to travel in the past 12 months?"

In this setup, there are 2 unknown parameters to be estimated: π_s =the proportion of people who have paid a bribe in past 12 months and π_n = the proportion of people who have travelled by bus in the past 12 months. As with the original procedure, the manager uses a randomising device to decide which question he/she answers.

The introduction of this second unknown parameter requires the use of 2 independent samples with different probabilities of selection ($p_2 = 1 - p_1$ and $p_1 \neq p_2$). The respective probabilities of observing a "yes" response in each sample is given by:

$$\lambda_i = p_i\pi_s + (1 - p_i)\pi_n \quad for \quad i = (1, 2). \tag{A.2.1.1}$$

Where i represents the sample. So for each respective sample:

$$\lambda_1 = p_1\pi_s + (1 - p_1)\pi_n \tag{A.2.1.2}$$

and

$$\lambda_2 = p_2\pi_s + (1 - p_2)\pi_n \tag{A.2.1.3}$$

Solving these equations to get an estimate of the prevalence of the sensitive act gives:

$$\hat{\pi}_s = \frac{\left(\hat{\lambda}_1(1 - p_2) - \hat{\lambda}_2(1 - p_1)\right)}{p_1 - p_2} \tag{A.2.1.4}$$

[100]

[100] This estimate has a variance of: $Var(\hat{\pi}_s) = \left(\frac{1}{(p_1-p_2)^2}\right) * \left(\frac{\lambda_1(1-\lambda_1)(1-p_2)^2}{n_1} + \frac{\lambda_2(1-\lambda_2)(1-p_1)^2}{n_2}\right)$

Alternatively, one can simplify this procedure and go back to using only one sample by using an unrelated non-sensitive question whose population distribution is already known. For example:

Question 2: "Is your birthday in January?"

has an population occurrence of $\pi_n = 31/365$, therefore equation (A.2.1.4) becomes

$$\hat{\pi}_s = \frac{\hat{\lambda} - (1-p)\pi_n}{p} \tag{A.2.1.5}$$

[101]

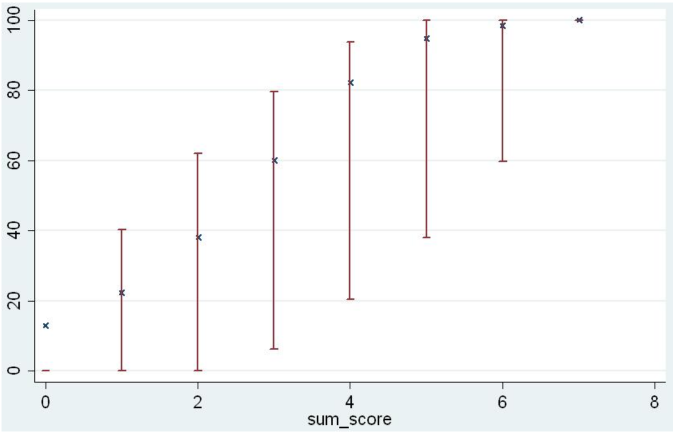
The two types of unrelated question RR designs differ in the second question that is asked. In the unrelated question RR with unknown population prevalence of the non-sensitive attribute, the distribution of the non-sensitive question is unknown, so two independent samples are required in order to estimate the sensitive trait. In the unrelated question RR with known population prevalence, the distribution of the non-sensitive question is known, so only one sample is required to estimate the occurrence of the sensitive trait.

Indirect Questions Ercolani [2006] uses data from the Labour Force Survey to study absence from work due to sickness. The author attempts to overcome four potential problems: identifying sickness rather than shirking behaviour; reducing the potential underreporting that might occur due to the stigma associated with being sick; generating measures that are consistent over the period being sampled; and overcoming the potential need to recall absences over long periods of time. The author tries to resolve these issues as follows: instead of asking about the number of days that the manager was sick, the interviewer uses answers to a question about usual hours worked per week versus actual hours worked in the week before the interview. Any difference in these two variables prompts a question concerning the reason for this difference somewhat later on in the survey; in the later question sickness is one of 13 response categories. Confidence in the survey responses is created by the fact that the survey is conducted away from the managers place of work and by an institution that is independent of the manager's employer. Despite this, potential issues are created when one considers that some workers choose to attend work even when sick; furthermore, any absences due to shirking work might be attributed to sickness due to the potential threat of admitting to shirking; finally, some employees might choose to use their holiday entitlement for sick leave. Therefore, in practice, the indirect measure of sickness might fail to be an accurate measure.

[101] Which has a variance of $Var(\hat{\pi}_s) = \frac{\lambda(1-\lambda)}{np^2}$

A.2.2 Appendix B

Figure 18: Sum Score Cumulative Distribution Function With Wald Binomial Confidence Intervals



Frequency (%) is measured on the vertical axis.

Figure 19: Maximum Likelihood Estimate Of Distribution of Responses Compared To The Actual Distribution And The Angels Assumption

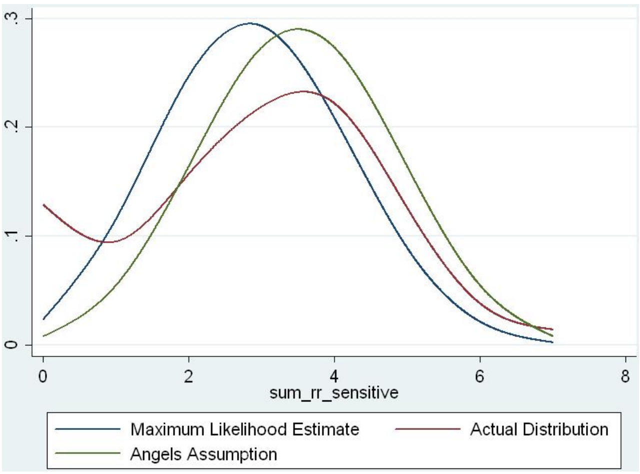


Table A.47: Logit Estimations

	(1) Estimation	(2) Marginal & Impact Effects	(3) Estimation	(4) Marginal & Impact Effects	(5) Estimation	(6) Marginal & Impact Effects
Sum Score: Q1-Q6	0.365***	0.082***				
	(0.025)	(0.006)				
Sum Score=1 (d)			1.551***	0.368***		
			(0.221)	(0.049)		
Sum Score=2 (d)			2.052***	0.471***		
			(0.203)	(0.040)		
Sum Score=3 (d)			2.254***	0.508***		
			(0.199)	(0.038)		
Sum Score=4 (d)			2.297***	0.518***		
			(0.203)	(0.037)		
Sum Score=5 (d)			2.380***	0.524***		
			(0.231)	(0.036)		
Sum Score=6 (d)			3.381***	0.616***		
			(0.337)	(0.025)		
Question 1 (d)					0.470***	0.106***
					(0.085)	(0.019)
Question 2 (d)					0.505***	0.115***
					(0.085)	(0.019)
Question 3 (d)					0.051	0.011
					(0.082)	(0.019)
Question 4 (d)					0.471***	0.107***
					(0.082)	(0.019)
Question 5 (d)					0.478***	0.109***
					(0.083)	(0.019)
Question 6 (d)					0.213***	0.048***
					(0.081)	(0.018)
owner_male (d)	0.128	0.029	0.110	0.024	0.134	0.030
	(0.112)	(0.025)	(0.114)	(0.024)	(0.113)	(0.025)
Age:Under 30 (d)	-0.310**	-0.067**	-0.317**	-0.067**	-0.310**	-0.068**
	(0.131)	(0.027)	(0.131)	(0.027)	(0.132)	(0.028)
Age:31-45 (d)	-0.070	-0.016	-0.064	-0.014	-0.069	-0.016
	(0.095)	(0.022)	(0.097)	(0.021)	(0.096)	(0.022)
Age:Over 55 (d)	-0.108	-0.024	-0.130	-0.028	-0.132	-0.029
	(0.139)	(0.031)	(0.139)	(0.030)	(0.140)	(0.031)
secondary (d)	-0.159*	-0.036*	-0.173**	-0.038**	-0.157*	-0.035*
	(0.082)	(0.018)	(0.083)	(0.018)	(0.082)	(0.019)
tertiary (d)	-0.240	-0.052	-0.268	-0.057	-0.221	-0.048
	(0.209)	(0.044)	(0.206)	(0.041)	(0.208)	(0.044)
manu (d)	0.030	0.007	0.074	0.016	0.017	0.004
	(0.094)	(0.021)	(0.095)	(0.021)	(0.094)	(0.021)
retail (d)	0.171	0.039	0.227*	0.051*	0.151	0.035
	(0.121)	(0.028)	(0.123)	(0.029)	(0.121)	(0.028)
medium (d)	0.102	0.023	0.079	0.018	0.087	0.020
	(0.104)	(0.024)	(0.104)	(0.023)	(0.104)	(0.024)
large (d)	0.857	0.209	0.781	0.188	0.789	0.192
	(0.600)	(0.148)	(0.630)	(0.157)	(0.618)	(0.154)
south (d)	-0.069	-0.016	-0.060	-0.013	-0.067	-0.015
	(0.082)	(0.019)	(0.083)	(0.018)	(0.083)	(0.019)
wave2 (d)	-0.001	-0.000	0.027	0.006	0.006	0.001
	(0.085)	(0.019)	(0.085)	(0.019)	(0.086)	(0.019)
Constant	-1.539***		-2.557***		-1.539***	
	(0.184)		(0.262)		(0.185)	
Pseudo R-squared	0.056		0.075		0.062	
Observations	3100		3100		3100	
Log-Likelihood	-1906.4		-1868.6		-1893.5	
Chi-Squared	232.5		210.4		250.9	
P Value	0.000		0.000		0.000	

$*p < 0.10, **p < 0.05, ***p < 0.01$

(d) for discrete change of dummy variable from 0 to 1. Dependent variable is the response to the last RR question (0=no; 1=yes).

Table A.48: Calculations Of Misclassification Errors

1	2	3	4	5	6	7	8	9	10	11	12	13
			Dichotomous Measures Of Reticence									
	Region	Obs.	Reticent(%) 7 + 3 Nos	Reticent(%) 7 + ≥2 Nos	Reticent(%) 7 + ≥1 No	Reticent(%) (7 Nos)	Reticent(%) (≥6 Nos)	Reticent(%) (≥5 Nos)	Reticent(%) (≥4 Nos)	Lower Bound Reticent (CKM)(%) (Based on question 10)	Alternative Lower Bound (%) (Based On Question 2)	Percentage of Reticent (Validation Data)
First Wave	Abia	92	4.3	6.5	6.5	6.5	23.9	43.5	62.0	13	6.5	34.4
	Abuja	94	3.2	4.3	4.3	4.3	12.8	29.8	53.2	10.6	2.1	28.7
	Anambra	100	12.0	21.0	21.0	21.0	29.0	49.0	60.0	32.0	14.0	39.6
	Bauchi	88	0.0	0.0	1.1	1.1	6.8	20.5	43.2	6.8	0.0	14.9
	Cross River	89	2.2	3.4	10.1	11.2	22.5	48.3	64.0	34.8	21.3	46.8
	Enugu	95	1.1	4.2	6.3	8.4	27.4	44.2	70.5	41.1	34.7	52.9
	Kaduna	92	1.1	2.2	3.3	4.3	19.6	39.1	62.0	30.4	19.6	42.2
	Kano	140	0.0	0.7	0.7	1.4	6.4	17.1	41.4	10.0	0.0	16.2
	Lagos	154	1.9	5.2	7.8	8.4	20.8	36.4	59.7	18.2	3.9	31.2
	Ogun	133	6.8	9.0	12.8	14.3	27.8	42.9	67.7	57.9	2.3	29.2
	Sokoto	61	1.6	1.6	1.6	1.6	4.9	23.0	52.5	21.3	11.5	33.7
Second Wave	Adamawa	83	6.0	7.2	9.6	10.8	13.3	21.7	56.6	13.3	6.0	27.1
	Akwa Ibom	78	3.8	5.1	12.8	30.8	35.9	48.7	62.8	38.5	33.3	48.0
	Bayelsa	86	3.5	9.3	9.3	10.5	22.1	40.7	61.6	30.2	14.0	38.1
	Benue	78	2.6	2.6	5.1	12.8	21.8	32.1	50.0	12.8	17.9	38.2
	Borno	86	11.6	14.0	15.1	16.3	26.7	39.5	65.1	27.9	11.6	35.3
	Delta	73	0.0	2.7	5.5	6.8	11.0	38.4	72.6	31.5	15.1	41.5
	Ebonyi	92	8.7	13.0	15.2	15.2	27.2	40.2	59.8	21.7	8.7	35.4
	Edo	44	6.8	9.1	11.4	11.4	15.9	31.8	56.8	13.6	0.0	15.6
	Ekiti	82	51.2	61.0	62.2	62.2	69.5	78.0	90.2	78.0	63.4	74.7
	Gombe	91	9.9	9.9	9.9	13.2	24.2	36.3	57.1	38.5	9.9	36.9
	Imo	82	2.4	3.7	14.6	19.5	40.2	52.4	74.4	36.6	39.0	53.3
	Jigawa	89	6.7	7.9	10.1	11.2	22.5	44.9	61.8	28.1	25.8	48.2
	Katsina	88	0.0	0.0	4.5	4.5	10.2	28.4	51.1	0	2.3	34.1
	Kebbi	96	2.1	2.1	2.1	2.1	6.3	27.1	47.9	0	18.8	45.3
	Kogi	87	37.9	41.4	43.7	43.7	59.8	70.1	83.9	72.4	63.2	74.1
	Kwara	76	6.6	7.9	7.9	7.9	11.8	38.2	56.6	31.6	21.1	43.9
	Nasarawa	84	52.4	54.8	56.0	56.0	65.5	75.0	83.3	64.3	57.1	70.5
	Niger	88	0.0	0.0	1.1	2.3	8.0	20.5	45.5	11.4	2.3	33.1
	Ondo	90	1.1	3.3	5.6	5.6	11.1	21.1	43.3	15.6	0.0	27.8
	Osun	82	2.4	2.4	3.7	3.7	13.4	25.6	48.8	9.8	2.4	33.6
	Oyo	55	7.3	10.9	10.9	14.5	34.5	63.6	87.3	70.9	63.6	75.6
	Plateau	62	24.2	27.4	30.6	32.3	48.4	54.8	71.0	54.8	38.7	57.8
	Rivers	50	0.0	0.0	0.0	0.0	2.0	16.0	40.0	0	4.0	26.2
	Taraba	66	3.0	4.5	4.5	4.5	10.6	22.7	59.1	30.3	0.0	23.6
	Yobe	81	4.9	7.4	7.4	7.4	13.6	34.6	53.1	35.8	1.2	29.2
	Zamfara	93	1.1	1.1	2.2	2.2	6.5	23.7	57.0	11.8	18.3	44.7
	National	3200	7.6	9.7	11.6	12.9	22.4	38.1	60.0	27.8	15.0	39.5
	First Wave	1,138	3.2	5.4	7.1	7.8	18.8	35.8	57.8	25.5	5.4	32.7
	Second Wave	2,062	10.0	12.0	14.0	15.8	24.4	39.4	61.2	29.1	20.3	43.3

Table A.49: Summary Statistics For Full Sample And Size, Sector, Region Groups

Group	Obs.	Mean	Std. Dev	Min	Max	Skewness	Kurtosis
1:Small; Manu; South	537	0.017	0.028	0	0.15	2.000	6.808
2:Small; Manu; North	877	0.031	0.042	0	0.15	1.150	3.082
3:Small; Retail; South	235	0.021	0.036	0	0.15	1.969	6.231
4:Small; Retail; North	218	0.022	0.035	0	0.15	1.633	4.927
5:Small; Other; South	314	0.021	0.036	0	0.15	2.076	6.740
6:Small; Other; North	354	0.032	0.041	0	0.15	1.334	3.981
7:Medium; Manu; South	190	0.013	0.024	0	0.1	2.079	6.914
8:Medium; Manu; North	150	0.042	0.044	0	0.15	0.650	2.166
9:Medium; Retail; South	39	0.015	0.021	0	0.06	1.071	2.431
10:Medium; Retail; North	42	0.035	0.045	0	0.1	0.650	1.546
11:Medium; Other; South	156	0.017	0.029	0	0.15	2.004	6.866
12:Medium; Other; North	77	0.049	0.041	0	0.1	0.050	1.402
13:Large; Manu; South	5	0	0	0	0	.	.
Full Sample	3200	0.025	0.037	0	0.15	1.500	4.329

A.3 Appendix For Chapter 3

Table A.50: Data Description

Category	Variable Name	Definition
Owner And Manager Characteristics	owner_male	Dummy=1 if owner is male
	owner_age	what is the age bracket of the sole owner or majority shareh
	owner_educ	what is the highest level of education... highest shareholde
	male_mgr	Dummy=1 if top manager is male
	mgr_age	what is the age bracket of the top manager?
	mgr_educ	what is the highest level of education of the top manager?
	mgr_experience	how many years of experience ... does the top manager have?
Nature Of Contracts (Variables Related To Trust)	sales_paid _before_delivery	% sales paid for before delivery?
	sales_paid _on_delivery	% sales paid for on delivery?
	sales_paid _after_delivery	% sales paid for after delivery?
	mat_paid _before_delivery	% materials paid for before delivery?
	mat_paid _on_delivery	% materials paid for on delivery?
	mat_paid _after_delivery	% materials paid for after delivery?
	orders_written	% purchase orders were written
	orders_oral_nowitness	% purchase orders were oral, without witness
	orders_oral_witness	% purchase orders were oral, with witness
	intermediate_sales	what % of total sales came from selling intermediate product
	principal_buyer	who was the principal buyer for establishment's output?
	primary_supplier	for how many years have you known the primary supplier of ma
	subcontract	in 2006, did you subcontract any part of your production?

Table A.51: Data Description

Variable Name	Definition
wave	Wave of questionnaire: 1 or 2
intcode	interviewer
super	supervisor
size:small	size==small (5-19 employees)
size:medium	size==medium (20-99 employees)
size:large	size==large (100 employees and more)
competitors_1	No Competitors
competitors_2	1 Competitor
competitors_3	2-5 Competitors
competitors_4	6+ Competitors

Table A.52: Data Description

Category	Variable Name	Definition
Randomised Response (RR) Questions	rr_personal_taxes	Have you ever paid less in personal taxes than you should have under the law?
	rr_business_taxes	Have you ever paid less in business taxes than you should have under the law?
	rr_job _app_misstatement	Have you ever made a misstatement on a job application?
	rr_office_phone	Have you ever used the office telephone for personal business?
	rr_promotion	Have you ever inappropriately promoted an employee for personal reasons?
	rr_not_pay	Have you ever deliberately not given your suppliers or clients what was due them
	rr_lie	Have you ever lied in your self interest?
	rr_hire	Have you ever inappropriately hired a staff member for personal reasons?
	rr_late	Have you ever been purposely late for work?
	rr_dismissal	Have you ever unfairly dismissed an employee for personal reasons?
RR Variables	understood_rr	Dummy=1 if manager understood how randomised response questions were working
	sum_rr_sensitive	Index (0-7) of the sum of 7 sensitive randomised response questions
	reticence	(Reticence) Dummy=1 if sum_rr_sensitive=0, 0 otherwise
	sum_rr_all	Index (0-10) of the sum of all 10 randomised response questions
	sum_rr_unsensitive	Index (0-3) of the sum of the unsensitive RR questions

Table A.53: Data Description

Variable Name	Definition
university	Dummy=1 if owner/majority shareholder has a university degree, 0 otherwise
validation_stax	Dummy=1 if firm underreports sales for tax purposes, 0 otherwise
validation_ltax	Dummy=1 if firm under-declares workforce (for tax purposes)
validation_tax	Dummy=1 if firm underreports sales, or workforce, for tax purposes

Table A.54: Data Description

Category	Variable Name	Definition
RR Variables (cont'd)	s1	1st sensitive question
	s2	2nd sensitive question
	s3	3rd sensitive question
	s4	4th sensitive question
	s5	5th sensitive question
	s6	6th sensitive question
	s7	7th sensitive question
	sum_ss6	Total Number of for 1st 6 sensitive questions

Table A.55: Data Description

Variable Name	Definition
secondary	Owner has secondary education
tertiary	Owner has tertiary education
manu	Dummy=1 if firm is in Manufacturing sector
retail	Dummy=1 if firm is in retail sector
wave2	Dummy=1 if wave=2
industry-location population	Number of firms (population) in this observation's industry-location cell

Table A.56: Data Description

Category	Variable Name	Definition
Owner And Manager Characteristics	oage_1	owner_age==30 years or less
	oage_2	owner_age==31-45
	oage_3	owner_age==46-55
	oage_4	owner_age==55 and more
	african	Dummy=1 if sole owner/majority shareholder is African, 0 otherwise
	indian	Dummy=1 if sole owner/majority shareholder is Indian, 0 otherwise
	lebaneseme	Dummy=1 if sole owner/majority shareholder is Lebanese or Middle Eastern, 0 othe
	oasian	Dummy=1 if sole owner/majority shareholder is from another Asian Country, 0 othe
	european	Dummy=1 if sole owner/majority shareholder is European/Caucasian, 0 otherwise
	other	Dummy=1 if sole owner/majority shareholder is from another ethic group, 0 otherw

Table A.57: Data Description

Variable Name	Definition
buyer_1	principal_buyer==your parent company or affiliated establishment
buyer_2	principal_buyer==large firms (more than 100 workers)
buyer_3	principal_buyer==medium private firms (20-100 workers)
buyer_4	principal_buyer==small private firms (less than 20 workers)
buyer_5	principal_buyer==individuals
buyer_6	principal_buyer==government or government agencies (including state-owned ent
buyer_7	principal_buyer==others
south	Dummy=1 if region==south

A.3.1 Appendix A

Table A.58: Summary Statistics

VARIABLES	N	mean	sd	min	max
owner_male	3,101	0.843	0.364	0	1
owner_age	3,100	2.350	0.870	1	4
owner_educ	3,097	4.821	1.830	1	11
male_mgr	492	0.882	0.323	0	1
mgr_age	492	1.884	0.748	1	4
mgr_educ	3,098	4.827	1.744	1	11
mgr_experience	3,197	11.49	7.046	1	45
sales_paid_before_delivery	3,200	33.66	33.63	0	100
sales_paid_on_delivery	3,200	54.34	36.88	0	100
sales_paid_after_delivery	3,200	12.00	20.94	0	100
orders_written	3,200	33.85	41.61	0	100
orders_oral_nowitness	3,200	38.35	42.09	0	100
orders_oral_witness	3,200	27.80	38.79	0	100
intermediate_sales	1,758	7.449	17.68	0	100
mat_paid_before_delivery	3,200	29.83	36.87	0	100
mat_paid_on_delivery	3,200	54.63	41.42	0	100
mat_paid_after_delivery	3,200	15.53	26.00	0	100
primary_supplier	3,174	7.148	6.211	0	151
subcontract	1,761	0.125	0.331	0	1
wave	3,200	1.644	0.479	1	2
intcode	2,062	307.4	143.3	3	509
super	2,062	34.83	24.54	1	205
size:small	3,200	0.792	0.406	0	1
size:medium	3,200	0.204	0.403	0	1
size:large	3,200	0.00344	0.0585	0	1

Table A.59: Summary Statistics (cont'd)

VARIABLES	N	mean	sd	min	max
competitors_1	1,763	0.0312	0.174	0	1
competitors_2	1,763	0.0159	0.125	0	1
competitors_3	1,763	0.185	0.389	0	1
competitors_4	1,763	0.767	0.423	0	1
rr_personal_taxes	3,200	0.503	0.500	0	1
rr_business_taxes	3,200	0.425	0.494	0	1
rr_job_app_misstatement	3,200	0.429	0.495	0	1
rr_office_phone	3,200	0.505	0.500	0	1
rr_promotion	3,200	0.404	0.491	0	1
rr_not_pay	3,200	0.379	0.485	0	1
rr_lie	3,200	0.518	0.500	0	1
rr_hire	3,200	0.407	0.491	0	1
rr_late	3,200	0.478	0.500	0	1
rr_dismissal	3,200	0.361	0.480	0	1
sum_rr_sensitive	3,200	2.909	1.727	0	7
reticence	3,200	0.129	0.336	0	1
sum_rr_all	3,200	4.411	2.100	0	10
sum_rr_unsensitive	3,200	1.502	0.958	0	3

Table A.60: Summary Statistics (cont'd)

VARIABLES	N	mean	sd	min	max
s1	3,200	0.503	0.500	0	1
s2	3,200	0.425	0.494	0	1
s3	3,200	0.429	0.495	0	1
s4	3,200	0.404	0.491	0	1
s5	3,200	0.379	0.485	0	1
s6	3,200	0.407	0.491	0	1
s7	3,200	0.361	0.480	0	1
sum_ss6	3,200	2.548	1.541	0	6
university	3,200	0.0388	0.193	0	1

Table A.61: Summary Statistics (cont'd)

VARIABLES	N	mean	sd	min	max
oage:≤30	3,100	0.154	0.361	0	1
oage: 31-45	3,100	0.452	0.498	0	1
oage:46-55	3,100	0.284	0.451	0	1
oage≥55	3,100	0.110	0.313	0	1
african	3,200	0.958	0.200	0	1
indian	3,200	0.00344	0.0585	0	1
lebaneseme	3,200	0.00156	0.0395	0	1
oasian	3,200	0.000313	0.0177	0	1
european	3,200	0.000313	0.0177	0	1
other	3,200	0.0362	0.187	0	1
buyer_1	3,179	0.00723	0.0848	0	1
buyer_2	3,179	0.00849	0.0918	0	1
buyer_3	3,179	0.0242	0.154	0	1
buyer_4	3,179	0.0780	0.268	0	1
buyer_5	3,179	0.862	0.345	0	1
buyer_6	3,179	0.00975	0.0983	0	1
buyer_7	3,179	0.0101	0.0998	0	1
south	3,200	0.462	0.499	0	1

Table A.62: Summary Statistics (cont'd)

VARIABLES	N	mean	sd	min	max
secondary	3,200	0.484	0.500	0	1
tertiary	3,200	0.0709	0.257	0	1
manu	3,200	0.551	0.497	0	1
retail	3,200	0.167	0.373	0	1
wave2	3,200	0.644	0.479	0	1
industry-location population	3,200	153.8	113.8	8	556

A.3.2 Appendix B

Table A.63: Tabulation Of Trust-Related Variables, By Reticence

		Percentage	
Variable	Categories	Possibly Candid	Reticent
Who was the **principal buyer** for this establishment's output?	Your parent company or affiliated establishments	0.61	1.46
	Large private firms (more than 100 workers)	0.72	1.70
	Medium private firms (20-100 workers)	2.20	3.89
	Small private firms (less than 20 workers)	7.33	10.95
	Individuals	87.10	80.29
	Government or government agencies (including state-owned enterprises)	0.94	1.22
	Others	1.08	0.49

Table A.64: Correlation Matrix

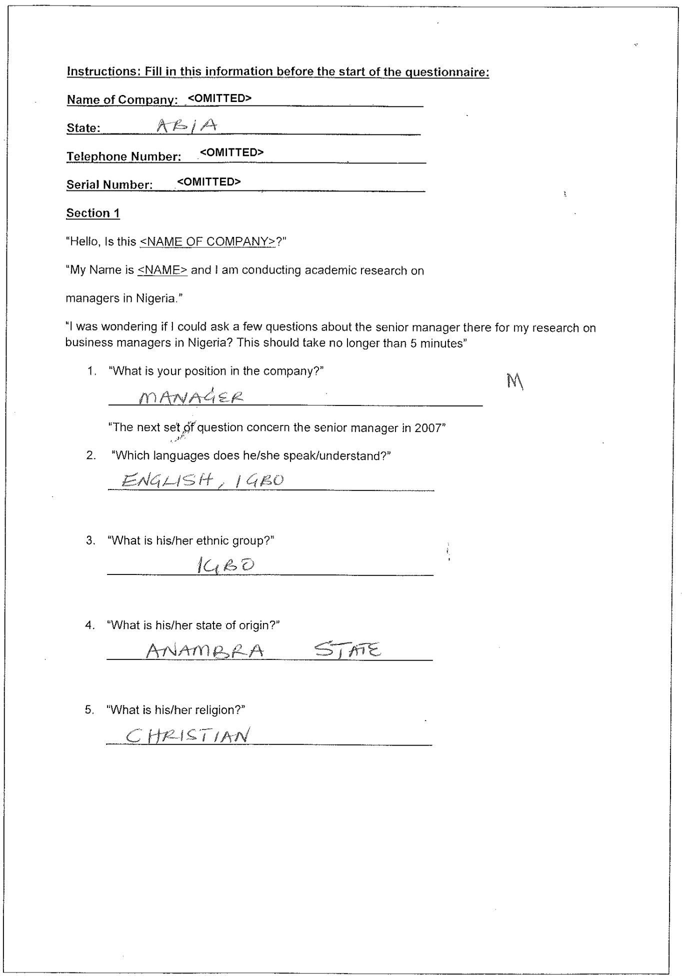
	(1)												
	r	mgr_experience	mgr_exp2	sales_paid_before_delivery	sales_paid_on_delivery	orders_written	orders_oral_nowitness	orders_oral_witness	subcontract	competitors_1	competitors_2	competitors_3	competitors_4
r	1												
mgr_experience	0.07***	1											
mgr_exp2	0.05**	1.0***	1										
sales_paid_before_delivery	0.03*	0.1***	0.08***	1									
sales_paid_on_delivery	-0.04*	-0.1***	-0.08***	-0.8***	1								
orders_written	0.1***	-0.006	-0.008	0.2***	-0.2***	1							
orders_oral_nowitness	-0.01	-0.01	-0.03	-0.07***	0.06***	-0.6***	1						
orders_oral_witness	-0.10***	0.02	0.04*	-0.1***	0.2***	-0.5***	-0.5***	1					
subcontract	-0.04	-0.04	-0.04	0.04	-0.2***	0.03	0.07**	-0.09***	1				
competitors_1	0.1***	0.05*	0.05*	0.03	-0.08***	0.04	-0.04	-0.006	-0.009	1			
competitors_2	0.08***	0.02	0.01	0.02	-0.04	-0.003	0.02	-0.02	-0.02	-0.02	1		
competitors_3	0.07**	-0.005	-0.005	0.03	-0.007	-0.04	-0.02	0.06**	-0.06*	-0.09***	-0.06*	1	
competitors_4	-0.1***	-0.02	-0.02	-0.04	0.05*	0.02	0.03	-0.05*	0.06**	-0.3***	-0.2***	-0.9***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.4 Appendix For Chapter 4

A.4.1 Appendix 1

Figure 20: Example Of Questionnaire Used For The Survey Of Managers



Instructions: Fill in this information before the start of the questionnaire:

Name of Company: <OMITTED>

State: ABIA

Telephone Number: <OMITTED>

Serial Number: <OMITTED>

Section 1

"Hello, Is this <NAME OF COMPANY>?"

"My Name is <NAME> and I am conducting academic research on managers in Nigeria."

"I was wondering if I could ask a few questions about the senior manager there for my research on business managers in Nigeria? This should take no longer than 5 minutes"

1. "What is your position in the company?"

   MANAGER M

   "The next set of question concern the senior manager in 2007"

2. "Which languages does he/she speak/understand?"

   ENGLISH, IGBO

3. "What is his/her ethnic group?"

   IGBO

4. "What is his/her state of origin?"

   ANAMBRA STATE

5. "What is his/her religion?"

   CHRISTIAN

A.4.2 Appendix 2

Further Estimations - Introduction A second stage analysis is performed to investigate the reporting of crime amongst firms that have been a victim of some sort of crime. Findings revel that, amongst the crime victims, bribe paying firms are more likely to refrain from reporting the crime to the police. This study suggests that corrupt firms do this because they wish to hide their malpractices from detection. These results indicate that the level of crime is probably underreported in areas with higher levels of corruption.

Further Estimations - Methodology If Similar-Ethnicity is a significant determinant of bribe payments then it might be useful to use this as an instrument for bribery in a two-stage model; since bribery can be endogenous to many variables, Similar Ethnicity can be used as an instrument in a first stage estimation in order to see the effects of bribery on firm outcomes. The current investigation seeks to assess whether bribe paying firms are less likely to report crime to the police when they have been a victim of crime. This exercise consists of: an estimation to model whether or not a firm pays a bribe; and an estimation to model whether or not a firm reports a crime, given that it is a victim of crime and pays a bribe. However, in order to allow flexibility into the estimation strategy, it might be useful to model the propensity to bribe and the propensity to be a victim of crime simultaneously. This is because one might raise the criticism that the errors from these two models might be correlated; i.e. that firms located in areas with high levels of crime might be more likely to be a victim of crime and this might affect their propensity to commit a crime (e.g. pay a bribe). Or firms might feel that they need to pay the police (or the police might demand bribes) in order to give them added protection from the crime in the area. On the other hand, firms that are located in areas with a relatively high police presence might be less likely to be a victim of crime but might also be more likely to be extorted by the police for bribes.

These ideas suggest that the errors determining the payment of bribes and being a victim of crime might be correlated. It might also be the case that the unobservables which determine whether one is a victim of crime affect ones propensity to report a crime. An area with a large mob might have higher crime rates but firm owners might fear that if they report the crime against them, then the mob will retaliate. This would be associated with a negative correlation between the errors in the victim and reporting equations. So, jointly considering the possible simultaneity between the payment of bribes and being on the receiving end of criminal activity; and the potential selectivity in the experiencing of criminal activity and the reporting of criminal activity, it seems plausible to firstly model the payment of bribes and being a victim of crime simultaneously, then model the reporting of crime as a second stage, controlling for potential selection in the reporting of crime.

To do this, this study first runs a bivariate probit model to consider the two binary outcomes of: paying a bribe (y_1); being a victim of crime (y_2). The investigation models the two outcomes as being determined by two unobserved latent variables:

$$\begin{aligned} y_1^* &= x_1^{'}\beta_1 + \epsilon_1 \\ y_2^* &= x_2^{'}\beta_2 + \epsilon_2 \end{aligned} \tag{A.4.2.1}$$

where ϵ_1 and ϵ_2 are jointly normally distributed errors both having a zero mean and a unit variance; their correlation is represented by ρ. The binary outcomes are observed under the following conditions:

$$y_1 = \begin{cases} 1 & \text{if } y_1^* > 0 \\ 0 & \text{if } y_1^* \leq 0 \end{cases}, \text{ and } y_2 = \begin{cases} 1 & \text{if } y_2^* > 0 \\ 0 & \text{if } y_2^* \leq 0 \end{cases},$$

with 4 possible outcomes: y_{00}, y_{11}, y_{10} and y_{01}.

Fitted values for $Pr(y_1 = 1|x)$ are used in a second stage estimation that models the propensity to underreport a crime on the payment of bribes. This estimation is run on a reduced sample of firms for whom $y_2 = 1$:

$$Pr(\text{At least one crime unreported}) = f(br\hat{i}be, x_i^{'}) \tag{A.4.2.2}$$

where the probability of underreporting crime is determined by a latent variable y_3^*:

$$y_3^* = \begin{cases} x_3^{'}\beta_3 + \epsilon_3 & \text{if } y_2 = 1 \\ - & \text{if } y_2 = 0 \end{cases},$$

Using these three models to investigate ethnicity, bribe payments and the underreporting of crime allows this study to test for correlation between ϵ_1 and ϵ_2; and between ϵ_2 and ϵ_3.

Further Estimations - Results This investigation now uses the Similar-Ethnicity variable as an instrument for bribery to see if this can be used to explain other economic outcomes for the firm. Table A.65 shows results from a bivariate probit estimation of the determinants of bribery and being a victim of crime. The coefficients on the model for bribery do not significantly change when estimating the model jointly as a bivariate model. Similar-Ethnicity is negatively related to being a victim of crime, but not significantly so. Nevertheless, the correlation between the errors of the two equations is positive (0.165) and significant at the 1% level, suggesting that firm-level corruption tends to occur in areas where other crimes against companies are also prevalent.

Fitted values from the bribery, (y_1), estimation are used in Table A.65 to estimate the effect of bribery on the propensity to report crimes. This is done in the last four columns of Table A.65. The models in this table are estimated only on firms that were victims of criminal activity. Results show a positive effect of bribery on a firm not reporting a crime. This suggests that corrupt firms are likely to underreport criminal activity performed against them, perhaps because they do not want the investigative authorities to discover any criminal acts that they have performed; or any illegal practices within their company. These results suggest that it is plausible to believe that the downward bias in the reporting of business malpractice is stronger than the upward bias. However, this result is only significant when clustering the standard errors by Similar-Ethnicity. When clustering the standard errors by zone, Similar-Ethnicity-zone, or using robust standard errors, the coefficient on the fitted values is not significant at the 10% level.

Table A.65: Second Stage Models Of Bribery And Crime On Matching, With Zonal Dummies

Model:	Bivariate Probit With Sample Selection		Robust Standard Errors	Clusterered S.E (Similar Ethnicity)	Clusterered S.E (Zone)	Clusterered S.E. (Similar Ethnicity
Dependent Variable	Unreported Crime	Victim of Crime	Unreported Crime	Unreported Crime	Unreported Crime	Unreported Crime
Pr(bribêdummy=1)	0.889		2.014	2.014***	2.014	2.014
	(1.850)		(1.903)	(0.177)	(2.261)	(1.718)
Similar Ethnicity		-0.083				
		(0.079)				
Zone==NC	-0.227	0.247**	-0.020	-0.020	-0.020	-0.020
	(0.226)	(0.114)	(0.243)	(0.105)	(0.211)	(0.165)
Zone==NE	0.688	0.118	1.170**	1.170***	1.170**	1.170***
	(0.553)	(0.141)	(0.484)	(0.088)	(0.501)	(0.383)
Zone==NW	0.204	0.240**	0.600	0.600**	0.600	0.600
	(0.447)	(0.113)	(0.397)	(0.290)	(0.432)	(0.432)
Zone==SE	0.120	0.287***	0.441	0.441**	0.441*	0.441*
	(0.309)	(0.110)	(0.281)	(0.176)	(0.265)	(0.267)
Zone==SS	-0.169	0.417***	0.228	0.228**	0.228	0.228
	(0.385)	(0.115)	(0.372)	(0.092)	(0.396)	(0.305)
Constant	0.538	-0.253***	-0.628	-0.628***	-0.628	-0.628
	(0.864)	(0.081)	(0.743)	(0.030)	(0.874)	(0.662)
Pseudo R-squared			0.026	0.026	0.026	0.026
Observations	571	1267	571	571	571	571
Chi-Squared	17.40		18.94	.	.	140.94
P-value	0.008		0.004	.	.	0.000
Rho	-0.790					
P-value	0.2023					

$*p < 0.10, **p < 0.05, ***p < 0.01$

Sample Selection Model: Heteroskedasticity Robust Standard Errors In Parenthesis. Similar Ethnicity: Dummy=1 if ethnicity of manager matches ethnicity of LGA Chair, 0 otherwise. Bribe Dummy: Dummy=1 if firm paid a bribe in the previous year, 0 otherwise. Victim Of Crime: Dummy=1 if firm was a victim on crime in the previous year, 0 otherwise. Unreported Crime: Dummy=1 if at least one crime went unreported, 0 otherwise. Unit of observation is the firm. Similar Ethnicity is calculated using manager data collected by the Author. All other variables are calculated using data from the NBS/EFCC survey. "Zone==..." is a dummy equal to 1 if the firm is located in said geo-political zone, 0 otherwise. Zones are defined as: NC= {Benue, Kogi, Kwara, Nassarawa, Niger, Plateau, FCT (Abuja)}; NE={Adamawa, Bauchi, Borno, Gombe, Taraba, Yobe}; NW={Jigawa, Kaduna, Kano, Katsina, Kebbi, Sokoto, Zamfara}; SE={Abia, Anambra, Ebonyi, Enugu, Imo}; SS={Akwa Ibom, Bayelsa, Cross River, Delta, Edo, Rivers} ; SW={Ekiti, Lagos, Ogun, Ondo, Osun, Oyo}. The omitted zone is SW.

Further Estimations - Robustness Checks In order to consider the possibility that the errors from the first stage victim model, ϵ_2, are correlated with the errors relating to the propensity to report crimes, a Bivariate Probit Model With Sample Selection is run in Table A.66. Results from the two stages of this model are shown in the first 2 columns of numbers in Table A.65. The coefficient of correlation between the errors of the two equations, ρ, is negative but not statistically significant. Thus, the sub-sample probit estimation seems like a plausible option to estimate the reporting model.

Table A.66: Firm Level Bivariate Probit Estimations Of Bribery And Crime On Matching, With Zonal Dummies

Dependent Variable:	Bribe Dummy	Victim of Crime
Similar Ethnicity	-0.194**	-0.080
	(0.084)	(0.078)
Zone==NC	-0.345***	0.246**
	(0.121)	(0.114)
Zone==NE	-0.747***	0.117
	(0.163)	(0.142)
Zone==NW	-0.582***	0.242**
	(0.125)	(0.113)
Zone==SE	-0.295***	0.284***
	(0.114)	(0.109)
Zone==SS	-0.571***	0.416***
	(0.127)	(0.115)
Constant	-0.147*	-0.254***
	(0.084)	(0.081)
Observations	1267	
Chi-Squared	69.35	
P-value	0.000	
Rho	0.165	
P-value	0.001	

$*p < 0.10, **p < 0.05, ***p < 0.01$

Heteroskedasticity Robust Standard Errors In Parenthesis. Similar Ethnicity: Dummy=1 if ethnicity of manager matches ethnicity of LGA Chair, 0 otherwise. Bribe Dummy: Dummy=1 if firm paid a bribe in the previous year, 0 otherwise. Victim Of Crime: Dummy=1 if firm was a victim on crime in the previous year, 0 otherwise. Unit of observation is the firm. Similar Ethnicity is calculated using manager data collected by the Author. All other variables are calculated using data from the NBS/EFCC survey. "Zone==..." is a dummy equal to 1 if the firm is located in said geo-political zone, 0 otherwise. Zones are defined as: NC= {Benue, Kogi, Kwara, Nassarawa, Niger, Plateau, FCT (Abuja)}; NE={Adamawa, Bauchi, Borno, Gombe, Taraba, Yobe}; NW={Jigawa, Kaduna, Kano, Katsina, Kebbi, Sokoto, Zamfara}; SE={Abia, Anambra, Ebonyi, Enugu, Imo}; SS={Akwa Ibom, Bayelsa, Cross River, Delta, Edo, Rivers} ; SW={Ekiti, Lagos, Ogun, Ondo, Osun, Oyo}. The omitted zone is SW.

Further Estimations - Conclusions Amongst firms who have been the victim of crime, bribe paying firms tend to underreport the crimes committed against them to the police. The proposed reason was because of a fear that their business malpractices might be discovered. This result suggests that crime might be underreported in areas with higher levels of corruption. Future research could investigate whether the nature of underreporting differs amongst different crimes, conditional on being a victim of those crimes.

A.4.3 Appendix 3

This section documents the results from running aggregated estimations of the effect of ethnic; religious; and linguistic diversity on corruption at the local government and state levels. Different measures of corruption are created using the conceptual framework of Méndez & Sepúlveda [2010]. Weighted and unweighed measures of diversity are also created using the methodology first outlined in Greenberg [1956].

Ethnicity And Economic Outcomes The early papers of Mauro [1995] and Easterly & Levine [1997] found some evidence for a negative association between ethnolinguistic fractionalisation and economic performance, however, since these studies there has been relatively little success in showing a negative direct effect of fractionalisation on growth. Some authors argue for a conditional effect of ELF [Easterly , 2001], i.e. that ELF is bad for growth only in economies with sufficiently poor institutions. In support of this idea, Bluedorn [2001] and Alesina, Devleeschauwer, Easterly, Kurlat & Wacziarg [2003] show evidence that the negative relationship between ELF and growth is more stronger in more autocratic countries. Posner [2004] suggests that the effect of ELF on economic activity is only strong when one excludes politically irrelevant ethnic groups. In other words, the negative effect of diversity on growth is only supported by a restricted diversity index.

The economic literature has also studied the relationship between religious diversity; autocracy; and economic outcomes. Collier & Hoeffler [2002] find no causal effect of religious diversity on the risk of conflict. Alesina, Devleeschauwer, Easterly, Kurlat & Wacziarg [2003] argue for a negative effect of ethnic and linguistic fractionalisation on the quality of government, but for a zero effect of religious fractionalisation. The authors also use the regression of Easterly & Levine [1997] to provide evidence for a zero effect of religious diversity on growth. The result that seems to be coming out of this strand of literature is that religious fractionalisation has no causal effect on the quality of government or economic performance.

Later research [Montalvo & Reynal-Querol , 2005b] argues that there is a weak direct effect of ELF on growth and that the literature should study the effects of ethnic polarisation, instead of ethnic fractionalisation, on economic performance. In summary, whilst the early literature argued in favour of a direct effect of ELF on growth, this has been challenged and the later literature has turned to finding out the methods by whuch diversity might affect economic performance. The current work contributes to this literature by using the traditionally used measure of diversity, ELF, and complementing this by adding other measures of diversity, including polarisation, doing this should help to identify which of these channels, if any, is the main factor leading to an effect of ethnic diversity on economic outcomes.

More recent research has investigated ethnic and hometown favouritism amongst politicians [Burgess, Jedwab, Miguel, Morjaria & i Miquel , 2011, Do, Nguyen & Tran , 2013]. Both studies suggest that more public funds are directed to the hometown of the politicians in power and to regions with a significant population of people who share the same ethnicity as the politicians in power. The current study builds on this work by including individual level information about the etnicities of the agents within the economy. By using the ethnicity of the firm managers this study is able to relax the assumptions that: a) all people within the hometown of a given politicians are treated equally; and b) all people that live in an area dominated by the ethnic group of the local politician are treated equally. Relaxing these two assumptions might allow for this study to be more in line with the literature on ethnic favouritism, which seems to emphasise the ethnicity of a person rather than where the person lives. Relaxing these two assumptions allows for people living in the same local area but of different ethnicities to be treated differently. For example, people in Lagos state might have access to better public services (such as roads and hospitals) than people in Ebonyi state, however, the treatment that people receive when accessing those public services might vary depending on their ethnic group. People from regions other than Lagos might receive a poorer service than people originally from Lagos state.

The Potential Endogeneity Of Ethnic Diversity Most of the abovementioned literature implicitly assumes that ethnic groups are mutually exclusive objective categories that individuals can be classified into. Moreover, it assumes that this classification is commonly shared amongst members of populations, and that this classification is exogenous. These assumptions can be questioned. Firstly, it might not be a priori clear what the relevant ethnic groups are; i.e. the boundaries of the ethnic groups that people are supposed to classify others into might not be objectively known to all members of all groups. A second point is that, under most definitions and measures of ethnic diversity, it is usually the case that these measures are determined by economic choices or policies. In other words, ethnic diversity might not be exogenous but can be determined within the system. The freeing up of national borders in order to allow the free movement of labour in and out of a country is likely to affect the ethnic mix of that country, immigrants might move in to work and indigenes might migrate away from the country for the same reason. However, the policy of liberalising one's borders is likely to be driven by economic considerations, therefore, whilst the ethnic composition of a country might have an impact on economic outcomes, the economy might drive certain policies to be made that affect the ethnic mix of the country. These can range from the extreme policies, e.g. ethnic cleansing[102], to the less extreme choices, e.g. creating costs/benefits for certain ethnic groups to remain in a location [Alesina & La Ferrara , 2005]. Moreover, economic policies can create incentives for the relocation/migration of people of certain ethnicities into certain area, with the effect that the ethnic mix of these areas are adjusted. In Nigeria, the National Youth Service Corps., which requires graduates of Nigerian universities to serve for one year in a location within the country that is away from their state of origin, can have this effect of changing the ethnic mix of a particular region. However, a policy does not necessarily have to relate to ethnicity in order to change the ethnic mix of a region. For example, many of the Fulani tribe are farmers and herders, thus any agricultural policy in Lagos which affects the trade of cattle is likely to lead to a change in the composition of the Fulani tribe in Lagos.

In Nigeria, much of the commercial activity occurs in Lagos, which is also one of the most ethnically diverse cities in the country. There is much inward migration due to the economic prospects of setting up a business within Lagos. The same can be said for Abuja (the capital city). These ideas tend to suggest that fractionalisation might be endogenous to economic activity; and that the ethnic mix of an economy can change over time, albeit at a relatively slow pace.

Ethnic Identity Previous research has assumed that ethnic groups can be easily identifiable and are independent of one another. Furthermore, these ethnic groups are assumed to be homogeneous units. Despite this, some recent literature has questioned this assumption. Horowitz [2001] and Humphreys, Posner & Weinstein [2002] report cases from Burundi; Sri Lanka; and Ethiopia, where it was tougher to distinguish members of different ethnicities despite conflict being purported to be on ethnic grounds. In these cases, the ability to fake an accent or to dress in different clothes made it sometimes impossible for one to identify somebody's ethnicity, even if they resided in the same locality.

Furthermore, self reported ethnicities might be dependent on economic/governmental policies. If questions of ethnicity are politically charged, and if this is common

[102] An example of this might be Germany during World War II.

knowledge, for example if the government favours (disfavours) a particular ethnic group, then it may be beneficial to report (or refuse to report) that one belongs to that group. Wilkinson [2002] reports two case studies, one from a Bohemian town where a third of self reported Germans in the 1910 Census chose to report themselves as Czech in the 1921 census in order to avoid discrimination. The second came from the state of Punjab, in India, where, in an attempt to stop a poposed partitioning of the state, many Hindu Punjabi speakers self-reported that they spoke Hindi. This led to a 20 percentage point fall in the self-reports of Punjabi speakers in the 1961 census .

Prominent Versus Negligible Ethnic Differences Ethnicity is often given as the reason behind many conflicts. However, when looking at economic outcomes, ethnicity does not always appear to be a good predictor. For example, Sub Saharan Africa is the most ethnically heterogeneous area in the world. However the actual occurence of conflicts is much lower than would be predicted by the level of ethnic fractionalisation. It thus appears that ethnicity might matter in some instances and not in others. Posner [2002] looks at the drawing of the boundaries of Malawi and Zambia, which disected the habitation of 2 ethnic groups: Chewas and Tumbukas, both of whose population was seperated so that 2/3 remained in Malawi and 1/3 was in Zambia. In Malawi, both groups represent a relatively large proportion of the population whilst in Zambia they do not. He notes that the two groups are political allies in Malawi, while hostile to each other in Zambia. His interpretation of this is that ethnic and cultural differences do not inherently shape economic outcomes, but the political and economic competition drives the salience of cultural ties.

Moreover, ethnic ties can become salient over time within the same institutional structure. An example of this occurring is in Somalia, which, before the civil war in 1991, was regarded as relatively ethnically homogeneous due to 85% of its population belonging to the Somali ethnic group. However, with the onset of the war, the relevant level of aggregation was shifted to the clan, therefore making the country more fractionalised [Alesina, Devleeschauwer, Easterly, Kurlat & Wacziarg , 2003]. The issue of which level of aggregation is relevant for the analysis of ethnic fractionalisation is discussed by Fearon [2003] and Desmet, Ortuño-Ortín, Wacziarg [2009].

Migration Finally, in spite of the previous discussion, another potential source of endogeneity is migration. Urbanisation and changes in economic growth have induced migration that is likely to have altered the ethnolinguistic fractionalisation of economies. Economic policies are likely to have a role in this as well.

The Measurement Of Ethnic Diversity: Determining The Aspect Of Diversity To Measure The method in which one is to categorise ethnic groups is of importance, but is accompanied by great difficulty; many judgement calls; and political influence. When individuals within a population can be different in the colour of their skin; their language; their place of birth; and their religion, the question of how to distinguish between different ethnicities becomes somewhat trickier. In some countries the relevant dimension might be language, in others it might be skin colour.

The data used in [Easterly & Levine , 1997] comes from the Atlas Narodov Mira [USSR , 1964] which was compiled based on data from 1960 describing the historical origins of linguistic groups. One potential flaw with this data is that linguistic diversity is not necessarily equivalent to ethnic diversity. An example of this potential disparity is found in most countries of Latin America: where a similar language is spoken by many different ethnic/racial groups. Such heterogeneity in the manifestation of diversity amongst different countries has been addressed by the literature in different ways [Alesina, Devleeschauwer, Easterly, Kurlat & Wacziarg , 2003, Fearon , 2003]. Fearon [2003] attempts to choose the most relevant cleavages in order to construct his index of fractionalisation whilst Alesina, Devleeschauwer, Easterly, Kurlat & Wacziarg [2003] use the breakdown of ethnicities as reported by the original sources. The first method has the advantage of being closer to the conceptual framework; the second method has the advantage of avoiding the potential error that can occur from making judgement calls. The correlation between the 2 measures is 0.76 [Alesina & La Ferrara , 2005], suggesting some similarities but not absolute convergence.

A second flaw with the Atlas data is in the way it categorises groups. Some groups, which operate as distinct political actors, are aggregated into one ethnic group. For example, the Acholi and Lango groups in Uganda are combined into a single group despite their documented history of political rivalry [Kasfir , 1976]. This grouping of political competitors also occurs with the Nyamwezi and Sukuma groups of Tanzania; and the Hutu and Tutsi groups of Rwanda and Burundi. The corollary of this critique is that some groups, whose distinction has no economic or political relevance, are listed under seperate categories [Posner , 2004, Alesina & La Ferrara , 2005].

The Measurement Of Ethnic Diversity:The Choice Of Index Much of the economic literature on ethnicity uses an index of ethnic fractionalisation as a measure of diversity:

$$ELF = 1 - \sum_i s_i^2 \tag{A.4.3.1}$$

where s_i is the proportion of group i in the population. This measure captures the probability that 2 randomly selected individuals from a population will not be from the same group. The index reaches a maximum value of 1 when every individual in a population belongs to a different group and a minimum of 0 when all members of a population are of the same group.

Despite the use of this measure, many authors have suggested that polarisaion, not fractionalisation, is the important measure of diversity that has an effect on economic outcomes. With this in mind, Garcia-Montalvo & Reynal-Querol [2002] proposed an index of polarisation:

$$POL = 1 - \sum_{i=1}^{N} \left(\frac{1/2 - s_i}{1/2} \right)^2 s_i \tag{A.4.3.2}$$

which reaches a maximum when the population comprises of two groups of equal size. Garcia-Montalvo & Reynal-Querol [2002] find that this measure of polarisation is correlated with ELF at low levels of ELF, but that this correlation declines as ELF increase; at high levels of ELF the two measures are negatively correlated. Some studies find that ethnic polarisation is a better explainer of economic outcomes [Garcia-Montalvo & Reynal-Querol , 2002] while others find that ELF is the preferred measure [Alesina, Devleeschauwer, Easterly, Kurlat & Wacziarg , 2003].

Something that is omitted from both measures but is still potentially useful is the distance between the groups of interest [Caselli & Coleman , 2002]. However, this is somwhat difficult to do in practice. One method, used to calculate linguistic diversity, is to use the linguistic distance between groups in the language tree [Greenberg , 1956, Desmet, Ortuño-Ortín, Wacziarg , 2009, Desmet, Ortuño-Ortín, Weber , 2009]. Another possible method is to use differences in average income levels between the groups [Aghion, Alesina & Trebbi , 2004]. However, this methodology remains relatively unattempted in the literature. The current study builds on this literature by using the language tree to construct measures of linguistic distances between languages.

Nepotism And Discrimination Nepotism refers to showing favouritism to one's own kin or group; it has usually been used to refer to such favouritism amongst bureaucrats in the public sector, however, it can also be used to describe the relationship between government employee and firm manager when describing behaviour when parties are informed about the other person's ethnic identity. Several studies look at nepotism as a type of discrimination [Brandts & Solà, 2010, Slonim & Garbarino , 2008, Belot & van de Ven , 2009]. Others make a distinction between nepotism and discrimination [Berg, Dickhaut & McCabe , 1995, Fershtman, Gneezy & Verboven , 2005].

Regionalism And Ethnicity In Politics In Nigeria One of the things that has existed in Nigeria's history and continues to do so within the political makeup of the country is the numerous inter-ethnic tensions which are driven by different perceptions and values and also by clashes of interest between the various groups that exist within the nation's borders. The most important tensions are the result of economic, geographical, political and religious differences that were mostly created along ethnic and regional lines within the twentieth century. They are not necessarily because of historic differences between ethnicities or dislike of other cultures and/or traditions.

At the start of the twentieth century, other than on isolated occasions, the dominant ethnic groups of Nigeria had relatively little contact with each other. Measures that were put into place during the colonial period created more incentives for the dominant groups to compete with each other for resources. Differences in the capacity of each group to exploit the opportunities provided by the modernisation process and nationhood was followed by differences in the rates of economic development, modernisation and education among the various groups. Some groups, notably those within the Northern states, preferred to stick to tradition rather than vie for key roles in the government and trade & industry. Other groups were more willing to seize such opportunities.

The differences in willingness to grasp new opportunities led some southern groups to achieve a disproportionate amount of wealth and power within Nigeria. The education and willingness of some tribes to migrate allowed them to win many high positions in the Northern states. These included positions in the local government councils. The differences between some of the southern groups were primarily the result of competitions for economic resources an political power. The strive to acquire a position permeated into state corporations, the civil service, and university faculty positions. Ethnic support was sought in these personal conflicts, a huge part of the reason for this is due to the fact that there were few divisions along the lines of class, religion, or economic group. Therefore, if the chairman of an organisation was a particular ethnic group, it was generally assumed that every job in that institution would go to a person of that same ethnic group. Similarly, if a government minister was from a particular tribe, it was automatically assumed that a predominant share of the Ministry posts and assignments would go to candidates from that same tribe.

Some minority groups also distinguished themselves from the three main ethnic groups. These sentiments were driven by a number of things, including: discrimination, history, tradition, and religious & cultural differences. However, the level of harmony between the three main groups and the many minority ethnic groups varied. A major landmark was achieved by the minorities in the Western Region with the creation of the Mid-Western Region which contained the non-Yoruba inhabitants [Ndaguba , 1995].

Methodology - State & Local Government Level This section spells out the methodology used in this study. It also explains the research questions that are addressed in the analysis.

Measuring Corruption When studying corruption one of the most important things to consider is how it is measured. Different authors use different measures to study the causes and effects of corruption. At the micro-level, an important question to ask is whether a business that makes 10 corrupt transactions per annum each involving ₦5,000 should be regarded as more, less, or equally corrupt as a business involved in 2 corrupt transactions per annum each involving ₦25,000. Does their behaviour makes them significantly different from one another from the point of view of the economics of corruption? The current investigation will use different measures of corruption in order to try to answer this sub-question.

The issue of the definition and measurement of corruption applied to a business is important for both theoretical and empirical work. Different measures have popped up in the literature, for example Reinikka & Svensson [2003] discuss the use of survey data to measure corruption whilst Svensson [2003] and Rand & Tarp [2010] use firm-level data to look at the behaviour of businesses with regards to bribery incidence; bribe level; and (in)formality. Olken [2007] uses the log of the ratio of reported expenditure to actual expenditure to measure corruption (missing expenditure) in Indonesian road projects.

There does not seem to be an a priori reason to choose one measure/definition instead of another. However, a problem arises when we ask whether or not results would have changed if we chose a different measure. Méndez & Sepúlveda [2010] address this problem theoretically by constructing three different measures of corruption which capture the general definition of corruption. Their results show that the effect of policy on corruption can depend on which measure one uses.

Specifically, the authors define 3 measures of corruption: corruption index (CI); relative corruption index (CRI); and total corruption rents (CR). These are defined as follows:

1. Corruption Index: Measures the number of times that a corrupt deal is observed within an economy.

2. Relative Corruption Index: Measures the number of times that a corrupt deal is observed relative to the number of deals within the economy.

3. Corruption Rents: Measures the total amount of bribes collected by public officials.

The current study builds on the work of Méndez & Sepúlveda [2010] by empirically looking at different measures of corruption using information from a business survey of companies in Nigeria. Doing this will help to see if the data is consistent with their proposed theory. However, it should be noted that their results suggest that the 3 measures of corruption might behave differently under certain conditions in response to policy, not that they will necessarily behave differently. Therefore, any results that show an empirical consistency between the 3 measures do not necessarily refute the work of Méndez & Sepúlveda [2010] but serves to justify the use of different measures of corruption that have been proposed in the literature. In measuring the effect of ethnicity on bribery; the current study uses all three measures to serve as a robustness check.

The previous analysis spelled out three distinct measures of corruption: corruption index (CI); relative corruption index (CRI); and corruption rents (CR). This study shall estimate two types of corruption index (CI). The first is the proportion of bribing firms within the economy; i.e. the proportion of firms that admitted to paying a bribe. So, if, in a LGA, 30 percent of firms admitted to paying any type of bribe, then that LGA will have a CI(1) of 0.3. The second measure of CI is the number of different purposes for which a bribe was paid. E.g. paying a bribe to: deal with traffic offences; deal with customs officials; deal with environmental regulations. A company which paid bribes in two of these situation will have a corruption index(2) of 2. The mean average of these values across a region (LGA or state) will form the second corruption index, CI(2).

To measure the relative corruption index (CRI), this study uses the ratio of the number of different purposes for which a bribe was paid to the total number of purposes for which the business could have paid a bribe. Managers gave details about the different types of government officials that they had met with and the different types of officials that they had bribed. The latter of these will form the numerator in the CRI and the former will form the denominator. So, if a manager met with 7 different types of government officials and paid bribes to 3 types; then that firm will be given a relative corruption index of $\frac{3}{7}$. The mean average of these values across a region (LGA or state) will form the relative corruption index, CRI.

An estimate of the revenue collected in bribes (CR) will be generated by adding all bribe reports by the companies in the sample. This will be analysed nationally, by state and by LGA. The log of these values will form the measure of corruption rents (CR).

Corruption And Ethnic Diversity ELF is defined as the probability that two randomly selected individuals from a population will not belong to the same ethnolinguistic group. Considering a country that has a population of N individuals who are distinguished by K ethnolinguistic groups, representing each group by $i = 1, ..., K$. Each individual belongs to one group only and the number of people in group i is represented by N_i. The current study allows for flexibility by not imposing any restrictions on the geographical location of any member of a group, therefore, individuals of ethnolinguistic group i can all live in the same region or be seperated and live in different regions. Since each individual belongs to only one group the sum of group populations will be equal to the size of the total population:

$$N = \sum_{i=1}^{K} N_i \tag{A.4.3.3}$$

The share of the population belonging to ethnolinguistic group i can be expressed as the ratio of the size of the ethnic group to the total population:

$$s_i = \frac{N_i}{N} \tag{A.4.3.4}$$

Therefore the sum of population shares will equal 1: $\sum_{i=1}^{K} s_i = 1$

ELF is defined as the probability that two randomly selected individuals will not belong to the same group. Formally:

$$\begin{aligned} ELF &= 1 - \sum_{i=1}^{K} s_i^2 \\ &= \sum_{i=1}^{K} s_i - \sum_{i=1}^{K} s_i^2 \\ &= \sum_{i=1}^{K} s_i - s_i^2 \\ &= \sum_{i=1}^{K} s_i \left(1 - s_i\right) \end{aligned} \tag{A.4.3.5}$$

The index of ELF satisfies the primary requirements of a diversity measure as stated by Shannon [1948]:

- For a fixed number of groups, the measure reaches a maximum when all group sizes are equal.
- If all groups sizes are equal, then the index increases with the number of groups in a society.

Much of the story surrounding corruption and ethnic fractionalisation stems from the seminal work of Shleifer & Vishny [1993]. Using an industrial organisation framework, they compare a situation where public officials collect bribes (set a bribe price) to that of an oligopolist setting the price for a good/service. The result being that bribe maximisation is occurs under similar conditions to profit maximisation in a collusive oligopoly, i.e. when an increase in the market bribe amount can be more easily detected and punished. An example of such a case might be a homogenous society where any movement in the bribe price from the normal amount might become known to the kin group.

In his seminal paper on corruption and growth, Mauro [1995] suggests that more heterogeneous societies countries will experience more corruption due to bureaucrats favouring their own ethnic group; in the context of the current study, this will mean bureaucrats collecting fewer or smaller bribes from their own ethnic group; and more or larger bribes from other ethnic groups. He uses an index of ethno-linguistic fractionalisation (ELF) as an instrument for corruption. More recent work has questioned the strength of identification in Mauro's paper. [Shaw, Katsaiti and Jurgilas , 2011] find Mauro's instruments to be "weak" as defined by the economics literature. The current study address this issue by looking at the link between bribery and ethnic fractionalisation within an economy. It focuses on the ethnic groups of business owners in relation to that of: local government chairmen and; the majority ethnic group of the state[103]. If the assumption is correct and ELF is a statistically strong determinant of corruption, then business owners whose ethnicity matches that of local government chiefs should be expected to pay fewer bribes than those who have a mismatch. Also, business owners whose ethnicity matches that of the majority (in the state) will have a lower propensity to bribe than those whose ethnicity do not match that of the state majority. This study tests the effect of ethnic fractionalisation on the amount of bribe that is paid. According to the ELF assumption [Shleifer & Vishny , 1993] regions with a higher ELF index should see a lower (mean) average bribe level and a higher incidence of bribery. Furthermore, regions with a lower ELF index (more homogenous states) should have a higher aggregate bribe payment but a lower incidence of bribery. The study thus adds to the literature on the causes of corruption and also that of the role of ethnicity and gender in African trade networks [Fafchamps , 2000, 2003]. The preceding ideas drive hypotheses 1 and 2.

Hypothesis 1 ***Test Of ELF As A Determinant Of Corruption (a):*** *Regions with a higher index of ELF will have a higher proportion of bribing firms*

[103] These are usually identical.

Hypothesis 2 ***Test Of ELF As A Determinant Of Corruption (b)**: Regions with a higher index of ELF will have a smaller amount of money collected in bribes*

It is important to note that these hypotheses only speak of the occurrence of bribery and the total amount of money collected in bribes, respectively; they do not mention the size of the average bribe payment amongst firms. Hypothesis 1 relates to the occurrence of bribery: CI; whereas Hypothesis 2 is about the total amount of money collected as bribe: CR. These hypotheses, if useful, provide a further justification for having different measures of corruption, each measure potentially captures something different, therefore allowing for different hypotheses to be tested.

This study follows much of the literature by using the probability that two randomly selected individuals are not of the same group as a measure of fractionalisation. With the current dataset this can be applicable to ethnicity; and religion. Using the same notation, this is defined as:

$$F = 1 - \sum_{i=1}^{K} s_i^2 \tag{A.4.3.6}$$

where s_i is the share of the population belonging to the ith group; and K is the total number of groups. Heterogeneity increases in F so that a value of 0 indicates a perfectly homogeneous group whilst a value of 1 indicates perfect heterogeneity.

This chapter uses similarity indices (See Section A.4.3) to weight different ethnicities according to how related they are. This measure is conducted using information from Barrett, Kurian & Johnson [2001]. Barrett, Kurian & Johnson [2001] classifies ethnocultural groups using a 6 character alphanumeric code that identifies race, geographic race, major culture area, stylized colour, ethnocultural family, and unique people. This is shown in table A.67.

Relative Similarity Between Ethnic Groups The previous analysis treats each ethnic group as being distinct. This is useful for trying to find out the level of diversity within an economy. One shortcoming of this approach is that it does not take into account the similarities between ethnicities. For example, the Ebira and Igbriran tribes are more closely related than the Ebira and Fulani tribes. To take another example, a Yoruba and a Hindu person are less ethnically similar than a Yoruba and a Hausa person; yet all these ethnicities are present within Nigeria and all are present within the current dataset. Not taking into account the heterogeneity in the (dis)similarity between ethnicities might be a source of bias in the construction of an index of fractionalisation. This is because two populations: one with a 50/50 mix between Yorubas and Igbos, the other with a 50/50 mix

Table A.67: Codes For The Classification Of Ethn-ocultural Groups

Race (biological grouping)	Geographical Race (continental race)	Major Culture Area (stylized colour)	Ethnocultural Family (local race)	People (subfamily)
A: Australoid	A: African	B: Black		
B: Capoid	E: European	G: Grey		
C: Caucasoid	F: Afro-American	N: Brown		
M: Mongoloid	I: American Indian	R: Red		
N: Negroid	L: Latin American	T: Tan		
	M: Middle Eastern	W: White		
	N: Indo-Iranian	Y: Yellow	01-71	a-z
	O: Oceanic			
	P: Pacific			
	R: Arctic Mongoloid			
	S: Asian			
	U: Austro-Asiatic			
	Y: Early African			

between Yorubas and Chinese; would receive the same index of fractionalisation, however, one might call the population consisting of a solely African population (Yoruba/Igbo) less fractionalised as the one consisting of an African and Asian population (Yoruba/Chinese). The current study takes the (dis)similarities between the ethnic groups into account by classifying them based on their ethnocultural groups.

Each ethnic group in the sample is coded in the same manner as the previous literature [Fearon , 2003], according to the codebook of the World Christian Encyclopedia [Barrett, Kurian & Johnson , 2001]. The methodology for this classification is displayed in Table A.67. Each ethnic group's code consists of a 5-6 alphanumeric code. The first 3 characters are letters, the next 2 are numbers and the sixth character , when present, is a letter. The first character,{A,B,C,M,N}, denotes the racial/biological grouping of the people within the ethnic group, either: Australoid, Capoid, Caucasoid, Mongoloid, or Negroid. The second character, {A,E,F,I,L,M,N,O,P,R,S,U,Y}, denotes the geographical/continental race of the people: African, European, Afro-American, American Indian, Latin American, Middle Eastern, Indo-Iranian, Oceanic, Pacific, Arctic Mongoloid, Asian, Austro-Asiatic, and Early African. The next character, {B,G,N,R,T,W,Y}, denotes the major cultural area of the group, or its stylized colour: Black, Grey, Brown, Red, Tan, White, and Yellow. The next 2 characters are numbers from 01 to 71 and denote the ethnocultural family, or local race, of the group. When there are different types of people within a local race a 6th character is used, from a to z, to denote this subfamily.

Similarity between any two ethnic groups is measured as the share of common code that they have. So, the Yoruba tribe (code: NAB59n) and the Urhobo tribe (code:NAB59z) will have a similarity index of 5/6. Furthermore, the Yoruba tribe and the English (code:CEW19i) will share a similarity index of 0/6. Different weights will be used to calculate dissimilarity measures (see Section A.4.3) to place greater emphasis on either on large differences (deep cleavages), e.g. the difference between the Yoruba and the English; or on smaller differences (finer distinctions), e.g. the difference between the Yoruba and the Urhobo [Desmet, Ortuño-Ortín, Wacziarg , 2009].

Fractionalisation Versus Dominance Recent literature has made a distinction between ethnic fractionalisation and ethnic dominance. Fractionalisation is the situation where society is divided into separate ethnic groups. Dominance, on the other hand, occurs when at least one minority group faces a majority. The argument goes that: whilst ethnic fragmentation might give rise to competing ethnic groups and an inability to co-operate, it is ethnic dominance that causes the victimisation of individuals [Collier , 2001]. In the context of the current study, both fractionalisation and dominance might have an impact on whether or not a randomly selected firm pays a bribe to a public official. This study argues that the distinction between the two is not significant in explaining the bribing behaviour of different ethnic groups. To the extent that factionalisation is more of a measure of heterogeneity, the index of ELF should suffice as a description of the ethnic makeup of the population of firm managers.

Ethnic fractionalisation (EF) increases as the number of ethnic groups increases, therefore, the more groups in a society, the more fractionalised it is, ceteris paribus. EF also increases as the groups become more equal in size, so a society that consists of different groups of the same size is more fractionalised than a group with the same number of groups but of unequal size.

In addition to the test of the ELF assumption that is inherent in Hypotheses 1 and 2, this study looks at the ways in which the polarisation of one or more ethnic groups might be related to bribery. In doing this, the study tests the effect of polarisation (or group dominance) on the propensity to bribe.

Hypothesis 3 ***Discrimination In The Propensity To Bribe****: Within each region, managers who belong to the dominant ethnic group will be less likely to pay bribes than managers who do not belong to the dominant ethnic group*

Hypothesis 4 ***Discrimination In The Amount Of Bribe Paid****: Within each region, bribe paying managers who belong to the dominant ethnic group, will , on average, pay less in bribes than the bribing managers who do not belong to the dominant ethnic group*

Hypotheses 3 and 4 can also be applied to groupings of individuals based on categories other than ethnicity. For example: migrant status[104]; religion; and gender. Doing this allows one to see whether discrimination along ethnic lines is stronger than religious or gender discrimination.

Polarisation In order to measure the degree of ethnic polarisation/dominance, this study uses the index of polarisation (RQ) proposed by Montalvo & Reynal-Querol [2005a]:

$$RQ = \sum_{i=1}^{K} s_i^2(1 - s_i) \tag{A.4.3.7}$$

The RQ measure of polarisation and the traditional measure of ELF both share the feature that they do not take into account distances between different ethnic groups. Therefore, the difference between Yoruba and Hausa is equivalent to the difference between Yoruba and English. The RQ index of polarisation reaches a maximum when there are two groups of the same size.

Corruption And Linguistic Diversity Previous literature that use ELF as a measure of diversity conflate two distinct concepts: ethnicity diversity and linguistic diversity. Measures of ELF taken from the Atlas Narodov Mira [USSR , 1964] are subject to this problem. This study addresses this problem by using seperate measures of ethnic and linguistic diversity. Traditional measures of linguistic diversity can be constructed by assigning each individual to a language group (usually their mother tongue) and using the same method as the calculation of ELF (equation (A.4.3.5)). This method, however, has two problems, which the current study seeks to address. The first problem is the fact that it treats all languages as being equally distant from one another. The second problem is that is does not allow for people to speak more than one language.

The current study follows Greenberg [1956], Desmet, Ortuño-Ortín, Wacziarg [2009] and Esteban, Mayoral & Ray [2012] and uses the linguistic distance between firm managers within the population to construct a measure of linguistic diversity. This allows for the linguistic fractionalisation index to be weighted by how closely related the languages are to one another.

By using the weighted versions of the linguistic diversity index, this investigation captures how related the languages are to each other, which provides more information than treating all languages and dialects as independent. Two languages from the Indo-European family of languages can be said to be more related

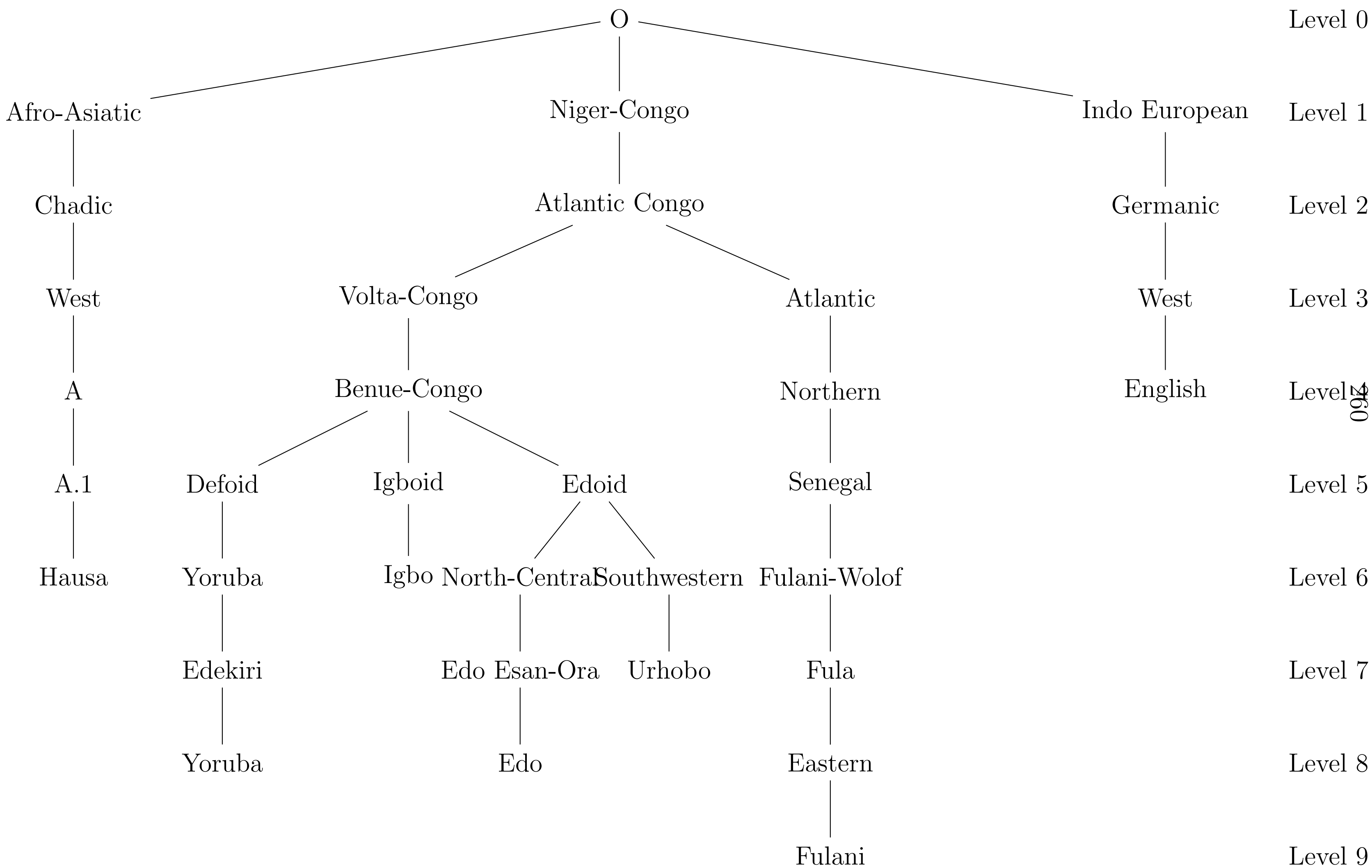
[104] An individual is defined as a migrant if their current location/state of residence is different from their state of origin. They are defined as a non-migrant if they originated from their current state of residence.

than a language from the Afro-Asiatic family and another from the Sino-Tibetan lineage. For example, French and Spanish are more closely related to each other than Arabic and Mandarin. With this in mind, a country that contained people who only spoke either French or Spanish would not be as diverse as a country whose members spoke Arabic or Mandarin since in the latter case the languages are much less related than in the former case.

An example of a typical language tree is shown in Figure 21. Each language, within a dataset, is found on the end branch within a tree. The weights are constructed as follows. The languages of the world can be organised into a language tree that captures their lineage, as shown in Figure 21. For example, all Afro-Asiatic languages will have a common subtree, with further splits creating more sub-subtrees all the way to the current map of languages. Modern research notes a potential maximum of 15 steps [Esteban, Mayoral & Ray , 2012]. Cultural diversity can be approximated by a lack of closeness on the language tree. One can define the proximity between 2 languages k and l, c_{kl}, as the ratio of the number of common branches to the largest number of branches for any language in the dataset, $c_{kl} \equiv \frac{n_{k=l}}{m}$. Where $0 < m \leq 15$ is the maximum possible number of branches. In the current study $m = 11$. If k and l are the same, then c is set to equal 1; and if the two language trees diverge at the first node, then c is set to equal 0. Using this definition, as in Fearon [2003] and Desmet, Ortuño-Ortín, Wacziarg [2009], the dissimilarity between the two languages can be defined as: $d_{kl} = 1 - c_{kl}^{\delta}$ for some $\delta > 0$. This measure of linguistic diversity looks at language families in addition to individual languages. Such an approach can provide more information due to the fact that some languages are more closely related to each other than to other languages; and that this closeness can also be reflected in a similarity of culture. A country that looks diverse when focusing on the number of different languages might cease to appear diverse when taking the family of languages into account.

A problem arises when the number of branches varies among different language families and subfamilies. For example, in Figure 21, the English language has 4 branches while Fulani has 9 branches. To rectify this problem the current study uses the methodology of Desmet, Ortuño-Ortín, Wacziarg [2009], who extend the tree by assuming that all modern languages are equally distant from the root, where distance is defined as the number of branches between the language and the root. In the current study, this will imply that 7 fictitious branches from levels 5 to level 11, will be added to the English language so that the total number of branches is equal for all languages. Desmet, Ortuño-Ortín, Wacziarg [2009] find that this method is not significantly different from other alternatives.

Figure 21: Phylogenetic Tree Of The Most Common Languages In The Nigeria Dataset



A natural problem to address is the choice of δ. Previous literature has used $\delta = 0.5$ [Fearon , 2003]; $\delta = 0.05$ [Desmet, Ortuño-Ortín, Wacziarg , 2009, Desmet, Ortuño-Ortín, Weber , 2009, Esteban, Mayoral & Ray , 2012]; and effectively $\delta = \infty$ [Montalvo & Reynal-Querol , 2005a]. The choice of δ is not trivial since it determines how much weight is given to linguistic dissimilarity. Low values of δ will distinguish languages with few branches in common; as δ increases , more weight is attached to small differences whilst larger differences are weighted proportionally less. In $lim_{\delta\to\infty}$ the smallest differences is indistinguishable from relatively deeper language cleavages. This study uses both $\delta = 0.05$ and $\delta = 0.5$.

The second problem that this study addresses is polylingualism. Standard measures of linguistic diversity assign individuals of a population to one group each; and compare the distances between the different groups [Desmet, Ortuño-Ortín, Wacziarg , 2009, Desmet, Ortuño-Ortín, Weber , 2009, Esteban, Mayoral & Ray , 2012, Bossert, D'Ambrosio & La Ferrara , 2011]. The problem with this is polylingualism, where an individual can speak more than one language. By assigning individuals to only one group some information might be lost. The current investigation uses measures of linguistic diversity proposed by Greenberg [1956] to account for polylingualism in the country. These are: the weighted split-personality method; and the unweighted split personality method.

These methods treat every bilingual individual as two seperate individuals; every trilingual person as three separate persons; and so on in that fashion. For the unweghted version, fractionalisation is calculated in the same way as ELF. The weighted version uses distance weights before calculating the index. This study is the first study, to the best of the author's knowledge, that allows for polylingualism in its measure of linguistic diversity. Therefore, this chapter solves two problems: polylingualism and related languages.

To calculate linguistic diversity, this study uses the measures proposed by Greenberg [1956] which take into account the possibility of polylingualism: the weighted split-personality method; and the unweighted split personality method. The weighted version controls for the degree of relatedness between two languages. These methods are described in turn.

In a standard model of linguistic diversity where each individual understands only one language, so $\sum_{i=1}^{K} s_i = 1$, the probability of randomly successively selecting a speaker of language i, and a speaker of language j is the product: $s_i s_j$. In this case, linguistic diversity is measured as:

$$LD = \sum_{i=1}^{K}\sum_{j=1}^{K} s_i s_j \tag{A.4.3.8}$$

In the weighted versions of the diversity indices, this product is multiplied by a measure of lingual distance,d_{ij} (Section A.4.3). Weighted linguistic diversity is calculated as:

$$LD_w = \sum_{i=1}^{K} \sum_{j=1}^{K} s_i s_j d_{ij} \tag{A.4.3.9}$$

Equation A.4.3.8 is equivalent to: $1-\sum_i^K s_i^2$; and Equation A.4.3.9 is equivalent to $1-\sum_{i=1}^K \sum_{j=1}^K s_i s_j c_{ij}^{\delta}$ where c_{ij}^{δ} is the measure of similarity (closeness) between languages i and j.

From this setup one can derive the result that the weighted measure of linguistic diversity, LD_w will be no greater than the unweighted measure, LD.

Since:

$$\begin{aligned} LD_w &= \sum_{i=1}^{K} \sum_{j=1}^{K} s_i s_j d_{ij} \\ &= \sum_{i=1}^{K} \sum_{j=1}^{K} s_i s_j (1 - c_{kl}^{\delta}) \end{aligned} \tag{A.4.3.10}$$

Recall that the measure of closeness between two languages, c, is greater than or equal to zero; and less than or equal to 1. Also, the weight attached to the degree of closeness, δ, is greater than zero:

$$\begin{aligned} 0 \leq c \leq 1; and \\ \delta > 0 \end{aligned} \tag{A.4.3.11}$$

Therefore,

$$0 \leq LD_w \leq \sum_{i=1}^{K} \sum_{j=1}^{K} s_i s_j \tag{A.4.3.12}$$

This result can be reasoned intuitively. Excluding weights means that all languages are treated as equally distant from one another. For the sake of analysis, one can define this distance as being equal to 1; moreover, the only difference between the unweighted and the weighted versions of the index is that the unweighted version uses Equation A.4.3.9 and sets $d_{ij} = 1$ (or $c_{kl} = 0$) for all i, j. In other words, an unweighted index treats two languages within the same subtree as entirely seperate languages. By including positive weights that represent the degree of closeness between two languages one allows for some languages to be closer than others, giving a less fractionalised picture of the economy than the unweighted version

Polylingualism The split personality method uses the same methodology shown in Equations A.4.3.8 and A.4.3.9. However, in this case the unit of observation changes from the individual to the instance of the language. In other words, every individual who speaks 2 languages is counted as 2 seperate people; and similarly for trilingual people and so on. In the weighted version, all multiples are weighted by their linugistic distance to one another. This means that both 'between' individual differences and 'within' individual differences are weighted.

Corruption And Diversity In order to formally test the effect of diversity on corruption, this study builds on the work of Oguzhan [2008]. The current study uses a similar methodology; but contributes to the literature by including linguistic diversity to the analysis.

As in the previous literature, the following models are estimated:

$$Corruption_s = \alpha_0 + \alpha_1 Polarisation_s + \alpha_2 X_s' + \epsilon_s \tag{A.4.3.13}$$

$$Corruption_s = \beta_0 + \beta_1 Fractionalisation_s + \beta_2 Fractionalisation_s^2 + \beta_3 X_s' + \xi_s \tag{A.4.3.14}$$

where $Corruption_s$ is the set of corruption measures, that were described in Section A.4.3, in state/LGA s. $Polarisation_s$ is the set of ethnic and religious polarisation measures (Section A.4.3); $Fractionalisation_s$ is the set of ethnic, religious, and linguistic fractionalisation measures (Section A.4.3); and X_s is a set of explanatory variables. The previous literature suggests that $\beta_1 > 0$.

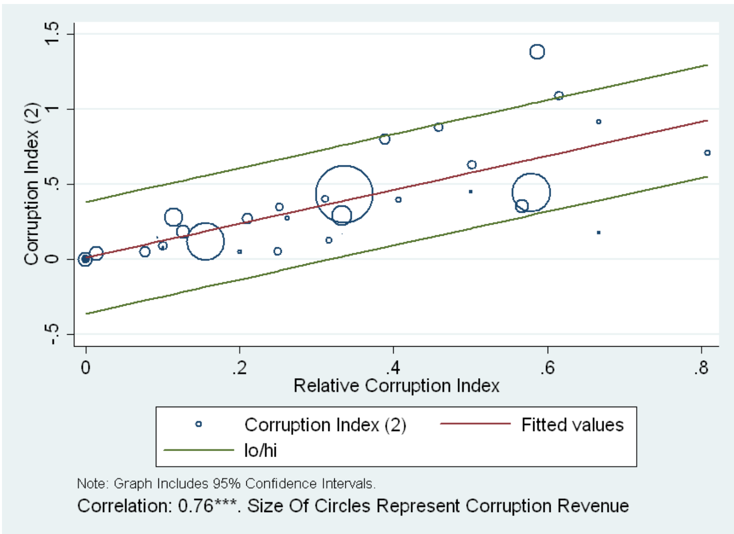
Results This section presents and discusses the results of the analysis on diversity and corruption. The impact of ethnic; religious; and linguistic diversity on corruption is presented via the methods described in the previous section. The analysis is conducted at the national level; regional level; state-level; local government level; and at the firm level.The raw data is collected at the firm level and aggregated up in order to run the estimations at the higher levels of aggregation. Data definitions are given in Tables A.68 and A.69 for the aggregated data and the firm level data, respectively. Summary statistics for the state and local government level data are shown in Tables A.70 and A.71, respectively. Summary statistics for the firm level data is shown in Table A.72.

Tables A.73 and A.74 have a combination of local government data and firm level data. Tables A.75 and A.76 show corruption measures and diversity measures, respectively, by region. Finally Figure 22 shows the relatioship between 3 of the main aggregated measures of corruption.

Tables A.77 to A.82 present the results from the estimations that are run at the state level of aggregation. Tables A.83 to A.88 show the results from the local government level estimations. Tables A.89 to A.92 show the results from the firm level estimations and Figure 15 provides information on how some of the geographical regions were aggregated into zones.

Results show no strong relationship between the state level measures of diversity and corruption, this applies to ethnic diversity, religious diversity and linguistic diversity. Also, this result remains when using either fractionalisation or polarisation as measures of diversity. This result also remains when using unweighted and weighted measures of linguistic diveristy.

Figure 22: Relationship Between The Main Measures Of Corruption



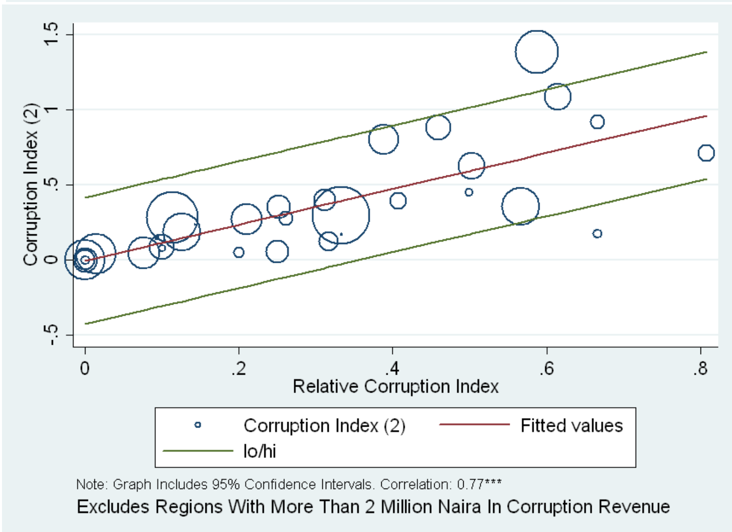


Table A.68: Data Definitions - Variables That Are Measured At The State Level And The Local Government Level

Category	Variable Name	Definition	Measurement
Measure Of Corruption	Corruption Index (1)	CI: Proportion Of Bribing Firms, By Region	$0 \leq x \leq 1$
	Corruption Index (2)	CI2: Average No. Of Types Of Bribes Transactions (Potential Max=11)	$0 \leq x \leq 12$
	Relative Corruption Index	CRI: Average 'Bribe Per Meeting', By Region	$0 \leq x \leq 1$
	Log (corruption rents)	log(Corruption Rents)	log(Naira)
Measures Of Diversity	Ethnic Polarisation	Ethnic Polarisation	$0 \leq x \leq 1$
	Religious Polarisation	Religious Polarisation	
	Linguistic Polarisation	Linguistic Polarisation	
	Ethnic Fractionalisation (Unweighted) (EF)	Ethnic Fractionalisation	
	Ethnic Fractionalisation ($\delta = .5$) (EF_{05})	Weighted Ethnic Fractionalisation	
	Ethnic Fractionalisation ($\delta = .05$) (EF_{005})	Weighted Ethnic Fractionalisation	
	Religious Fractionalisation (RF)	Religious Fractionalisation	
	Linguistic Fractionalisation (Unweighted) (LFU)	Unweighted Linguistic Fractionalisation	
	Linguistic Fractionalisation ($\delta = .5$) (LF_{05})	Weighted Linguistic Fractionalisation	
	Linguistic Fractionalisation ($\delta = .05$) (LF_{005})	Weighted Linguistic Fractionalisation	
	Linguistic Polarisation	Linguistic Polarisation Measure	
	EF Squared	EF Squared	
	RF Squared	RF Squared	
	LFU Squared	LFU Squared	
	Manufacture (Share)	Proportion of Firms in the manufacturing industry, by region	$0 \leq x \leq 1$
	Trade (share)	Proportion of firms engaging in international trade, by region	
	Avg. No. Of Languages Spoken	Number of languages that the manager understands (Max=5)	$1 \leq x \leq 5$

x represents the variable.

Table A.69: Data Definitions - Variables That Are Measured At The Firm Level

Category	Variable Name	Definition	Measurement
Dependent Variable	Bribe Dummy	Dummy=1 if firm admitted to bribing; 0 otherwise	{0;1}
Independent Variables	Similar Ethnicity	Dummy=1 if ethnicity of manager=ethnicity of local government chairperson, 0 otherwise	{0;1}
	Ethnic Mismatch	1-Similar Ethnicity	{0;1}
	Employee	number (category) of paid employees {Less Than 50; 50 to 100; 100 to 250; Over 250}	{1; 2; 3; 4}
	Trade	Dummy=1 if firm engages in international trade, 0 otherwise	{0;1}
	Foreign	Dummy=1 if foreign ownership $\geq$25%, 0 otherwise	{0;1}
	Security	Dummy=1 if company using security service, 0 otherwise	{0;1}
	Lage	Log of firm's age	

Table A.70: Summary Statistics For Variables Measured At The State Level

VARIABLES	N	mean	sd	min	max
Corruption Index (1)	37	0.211	0.135	0	0.582
Corruption Index (2)	37	0.320	0.338	0	1.381
Relative Corruption Index	35	0.299	0.226	0	0.808
Corruption Rents	37	541.9	1,332	0	7,180
Log (Corruption Rents)	35	4.838	1.951	-0.693	8.879
Ethnic Fractionalisation (EF)	37	0.407	0.276	0.0120	0.821
EF Squared	37	0.240	0.239	0.000143	0.674
Ethnic Fractionalisation ($\delta = .5$) (EF_{05})	37	0.070	0.052	0.003	0.199
EF_{05} Squared	37	0.008	0.009	0.000	0.040
Ethnic Fractionalisation ($\delta = .05$) (EF_{005})	37	0.009	0.011	0.000	0.068
EF_{005} Squared	37	0.000	0.001	0.000	0.005
Religious Fractionalisation (RF)	37	0.259	0.173	0.0185	0.507
RF Squared	37	0.0962	0.0945	0.000341	0.257
Linguistic Fractionalisation (Unweighted) (LFU)	37	0.650	0.0783	0.500	0.800
LFU Squared	37	0.428	0.101	0.250	0.640
Linguistic Fractionalisation ($\delta = .5$) (LF_{05})	37	0.599	0.0593	0.500	0.698
Linguistic Fractionalisation ($\delta = .05$) (LF_{005})	37	0.567	0.0647	0.482	0.669
Ethnic Polarisation	37	0.117	0.0562	0.00599	0.187
Religious Polarisation	37	0.126	0.0847	0.00917	0.249
Linguistic Polarisation	37	0.206	0.0271	0.139	0.250
Avg. No. Of Languages Spoken	37	2.324	0.219	2	2.800
Manufacture (Share)	37	0.135	0.115	0	0.615
Trade (Share)	37	0.0425	0.0452	0	0.165

Significant relationships between diversity and corruption begin to show up when running the estimations at the local government level. This applies to both ethnic and religious diversity, which both seem to have a relatively strong association with the proportion of corrupt firms and the log of the total amount of Naira paid (collected) in bribes. This result also applies to the firm level, where local government level averages of ethnic, religious and linguistic diversity are highly correlated with the individual firms propensity to bribe.

Relatively strong evidence is found for an ethnic/political network effect. A firm manager whose ethnicity matches that of his/her local government chairperson is less likely to pay a bribe than a manager of a different ethnicity to their local government chairperson. This result holds for the payment of bribes but not for the amount of bribe paid. No significant relationship was found between a shared ethnicity and the amount of bribe paid. Including the Similar-Ethnicity variable (a dummy equal to one if the manager's ethnicity is the same as that

Table A.71: Summary Statistics For Variables Measured At The Local Government Level

VARIABLES	N	mean	sd	min	max
Corruption Index (1)	350	0.198	0.298	0	1
Corruption Index (2)	350	0.228	0.609	0	5.500
Relative Corruption Index	200	0.262	0.337	0	1
Corruption Rents	350	57.48	234.6	0	3,000
Log (Corruption Rents)	152	3.376	1.817	-0.693	8.006
Ethnic Fractionalisation (EF)	350	0.320	0.354	0	1
EF Squared	350	0.228	0.329	0	1
Ethnic Fractionalisation ($\delta = .5$) (EF_{05})	350	0.028	0.044	0	0.207
EF_{05} Squared	350	0.003	0.006	0	0.043
Ethnic Fractionalisation ($\delta = .05$) (EF_{005})	350	0.003	0.006	0	0.072
EF_{005} Squared	350	0.000	0.000	0	0.005
Religious Fractionalisation (RF)	350	0.247	0.326	0	1
RF Squared	350	0.167	0.311	0	1
Linguistic Fractionalisation (LF)	292	0.583	0.0971	0	0.807
LF Squared	292	0.350	0.112	0	0.651
Linguistic Fractionalisation ($\delta = .5$) (LF_{05})	292	0.547	0.0722	0	0.720
Linguistic Fractionalisation ($\delta = .05$) (LF_{005})	292	0.527	0.0705	0	0.720
Ethnic Polarisation	350	0.0729	0.0886	0	0.250
Religious Polarisation	350	0.0651	0.0922	0	0.250
Linguistic Polarisation	292	0.228	0.028	0	0.250
Avg. No. Of Languages Spoken	292	2.272	0.393	1	4
Manufacture (Share)	350	0.137	0.267	0	1
Trade (Share)	350	0.0521	0.163	0	1

Language data is available for 1,267 firms. These 1,267 firms are located in 292 Local government areas. 198 local government areas had a total corruption rent of zero, hence, there are only 152 (350-198) observed values for the log of corruption rents.

Table A.72: Summary Statistics For Variables Measured At The Firm Level: Selected & Unselected Subsamples; And For Entire Sample

Variable	Statistic	Firms with Data on Ethnicity	Firms without Data on Ethnicity	All Firms	Test of Relationship (P and Z Values)
Bribe Dummy	Min	0	0	0	
	Max	1	1	1	
	Mean	.287	.242	.269	
	Median	0	0	0	0.022+
	Std. Dev.	.45	.43	.27	
Employee	Min	1	1	1	
	Max	4	4	4	
	Mean	1.53	2.18	1.79	
	Median	1	2	1	0.000+
	Std. Dev.	.95	1.29	1.15	
Trade	Min	0	0	0	
	Max	1	1	1	
	Mean	.07	.09	.08	
	Median	0	0	0	0.167+
	Std. Dev.	.26	.28	.26	
Foreign	Min	0	0	0	
	Max	1	1	1	
	Mean	.11	.17	.13	
	Median	0	0	0	0.000+
	Std. Dev.	.316	.372	.34	
Security	Min	0	0	0	
	Max	1	1	1	
	Mean	.99	.99	.99	
	Median	1	1	1	0.594+
	Std. Dev.	.08	.09	.08	
Age	Min	0	0	0	
	Max	90	90	90	
	Mean	16.3	20.26	17.9	
	Median	14	17	15	0.000−
	Std. Dev.	13.2	15.43	14.26	
Lage	Min	0	0	0	
	Max	4.50	4.50	4.50	
	Mean	2.45	2.68	2.54	
	Median	2.64	2.86	2.71	0.000−
	Std. Dev.	.94	.94	.94	
Similar Ethnicity	Min	0			
	Max	1			
	Mean	.627			
	Median	1			
	Std. Dev	0.484			
N		1,267*	843*	2,110*	

In each column, for each variable, the table reports the minimum, maximum, mean, median and standard deviation for each (sub)sample. The last row details the number of observations for each (sub)sample. *Only 1255, 834, and 2,089 observation for "Lage", respectively, due to log(0) being undefined. + P-Value was computed using a chi-squared test of independence between the variable and a "selection" variable equal to 1 if the managers ethnicity is known, 0 otherwise. The test on "employee" has 3 degrees of freedom, the tests on "bribe dummy", "trade", "foreign" and "security" each have 1 degree of freedom. − Z-Value calculated using a Wilcoxon-Mann-Whitney (rank-sum) test.

Table A.73: Summary Statistics Of Ethnicities Of Local Government Chairs: For Selected & Unselected Subsamples; And For Entire Sample

Ethnicity of Local Government Chairman	Selected For Analysis: Selected	Selected For Analysis: Not Selected	Total
Hausa	357	295	652
%	54.8	45.2	100.0
Igbo	224	112	336
%	66.7	33.3	100.0
Yoruba	471	273	744
%	63.3	36.7	100.0
Other	215	163	378
%	56.9	43.1	100.0
Total	1,267	843	2,110
%	60.0	40.0	100.0

Pearson χ^2_3 statistic = 18.626. P-value=0.000.

Table A.74: Cross Tabulation Of The Ethnicities Of Firm Managers And Local Government Chiefs

		Ethnicity Of Local Government Chief				
		Yoruba	Igbo	Hausa	Other	Total
Ethnicity Of Firm Manager	Yoruba	331	6	55	11	403
	(%)	26.12	0.47	4.34	0.87	31.81
	Igbo	93	192	53	35	373
	(%)	7.34	15.15	4.18	2.76	29.44
	Hausa	4	3	185	18	210
	(%)	0.32	0.24	14.60	1.42	16.57
	Other	43	23	64	151	281
	(%)	3.39	1.82	5.05	11.92	22.18
	Total	471	224	357	215	1,267
	(%)	37.17	17.68	28.18	16.97	100.00

Pearson χ^2_9 statistic = 1300. P-value=0.000. Percentages represent the percentage of the sample of 1,267 firms that exist within the respective cell. The main diagonal represents firm managers whose ethnicities match those of the local government chief. The off diagonal entries are firm managers who have a different ethnicity to that of the local government chiefs when splitting ethnicities into the 4 main groups of: Yoruba; Igbo; Hausa; and Other.

of the local government chair, and 0 otherwise) does not significantly alter the other independent variables in the model. This seems to suggest that the Similar-Ethnicity variable is picking up something new that was not previously within the data on firm level characteristics.

Also, when controlling for regional heterogeneity in the matching effect by including zonal dummies and interaction terms, the effect seems to appear only for the South-West region, which includes Lagos State, the commercial hub of

Nigeria and one of the most ethnically diverse regions in the country. This seems to suggest that much of the effect of ethnic networks on corruption might come from this region.

Summary Statistics

State Level Information Summary statistics for the variables that are measured or aggregated at the state level are given in Table A.70. The Corruption Index-1 (CI:1) represents the average number of bribing firms in each region. The mean value of 0.211 for the CI:1 indicates that the average proportion of bribing firms across all states was 0.21. The maximum value for any state is 0.582, which occurred in Lagos State. The minimum CI:1 was 0, which occurred in Jigawa and Yobe states. A skewness/kurtosis test for normality of this variable is unable to reject the null hypothesis of normality ($\chi^2_2 = 1.19$;P-velue=0.5528).

Corruption Index-2 (CI:2) represents the number of different types of bribe transactions that the firm engages in over the course of the previous year. The potential maximum number is 11 which corresponds to the 11 questions which asked about different types of bribe transaction. The actual range of values for the state level averages went from a minimum of 0 to a maximum value of 1.281, with a mean of 0.32 and a median of 0.24.

The relative corruption index measures the number of bribe transactions as a share of the total number of transactions that the firm engaged in (and was asked about) that had the potential for a bribe to take place. This variable has a potential range from 0 (representing the case where the firm met with public officials but did not report paying a bribe) to 1 (where the firm reported paying a bribe to every type of public official that it met with). The values in the data range from 0 to 0.808, with a mean of 0.299. A normality test on this variable also fails to reject the null of normality ($\chi^2_2 = 3.19$; P-value=0.2033). This variable has 2 missing values corresponding to the two states where no firms declared that they had met with a public official.

Finally, the corruption rents measures the total amount of bribes paid by firms within each state in '000 Naira. This is truncated at ₦0 (Yobe and Jigawa States) with a maximum of ₦7,180,000 (Lagos State), a mean of ₦542,000 and a median value of ₦151,000.

The three main linguistic (ethnic) fractionalisation measures reveal that weighting languages (ethnicities) by their similarity, represented by the number of common branches (code) that they share on the language tree (ethnicity code), reduces the measure of linguistic (ethnic) diversity in this case. The mean value of the state-level average number of languages spoken is 2.3, with a minimum of 2 and a maximum of 2.8. The null of normality cannot be rejected with this variable. An average of 13.5% of firms are in the manufacturing sector within each state; Kano State had the maximum value with 61% of its firms belonging to the manufacturing sector. 4.25% of firms engage in the direct importing or exporting of goods as part of their business, Lagos State had the maximum value for the share of its firms engaging in internationlal trade with 16% of them doing so.

Local Government Level Information Summary statistics for the variables that are measured or aggregated at the local government level are given in Table A.71. These variables are the same variables as described in Section A.4.3, the only difference is that they are constructed at the local government level. There are a total of 351 local government areas (LGAs) in the sample, this represents 45% of Nigeria's 774 Local Government Areas. Data on ethnicity exists for all 351 LGAs whilst data on language is only available for firms existing in 292 of the LGAs. Notable differences at this level of disaggregation include the maximum value for the total amount of Naira paid in bribes being ₦3 million. Also, the mean values for the diversity indices have generally reduced in magnitude whilst the maximum values have increased. Both results can be explained by the smaller cell size and lower level of aggregation when looking at the local government level, compared to the state-level. 198 local government areas contained firms who all reported paying zero bribes, therefore the log of corruption rents only exists for 152 observations.

Firm Level Information Summary statistics calculated at the firm level are displayed in Table A.72. This table shows summary statistics for the firms with data on ethnicity, firms without data on ethnicity and both groups combined. The following information relates to the 1,267 sub-sample of firms that have data on the ethnicity of their senior managers. Amongst these firms, 28.7% of them reported that they had paid a bribe. Information on the number of employees within each company was constructed using a categorical variable, the categories are: less than 50; 50-100; 100-250; and over 250. The median and mode values for the number of employees within the firm, amongst the sample of firms with ethnicity data, lie within the "less than 50" category. 7% of firms engage in international trade and 11% of firms had foreign owners whose share of ownership was greater than or equal to 25%. The majority of firms (99%) used a security service. The mean average age was 16 years, with a median of 14 years. Finally, 63% of the managers in this sub-sample belonged to the same ethnicity as the local government chairperson in the location of the company.

Local Government And Firm Level Information Table A.73 divides the data into 2 samples. Firms with information on the manager's ethnicity (Selected) and firms without information on ethnicity (Not Selected). The table cross tabulates this data along with the ethnicity of the local government chairperson of the local government area that the firm is located in. Ethnicity is divided into four categories: Hausa, Igbo, Yoruba, and Other. These categories were chosen because they are the most common ethnicities within the firm-level dataset; the individual level dataset and within Nigeria.

Using this information, the table shows the degree to which the respective ethnicities are selected into the sample. The data shows that firms in Hausa-headed LGAs are approximately equally selected as those in "Other"-Ethnic Group Headed LGAs, with 55-57% of the firms in each category being selected into the sample. Also, firms in Igbo headed LGAs are roughly equally selected into the sample as firms with Yoruba-headed LGAs, with 63-67% of the firms in each category being selected into the sample.

Table A.74 shows a cross-tabulation of the ethnicities of the local government chair and the ethnicity of the firm manager, for the selected sub-sample of firms that has information on the firm manager's ethnicity. The table shows the number of firms existing in each cell and also the percentage of the total sample that exists within that cell. Numbers along the main diagonal represent firms whose manager has the same ethnicity as the local politician. Conversely, firms existing in cells on the off diagonal represent firms whose managers have a different ethnicity to their local government chairperson. The firms existing in the main diagonal of this table comprise 67.79% of the sample of 1,267 firms. This is higher than the 63% of firms for whom it was previously stated that their manager's ethnicity matched that of the local government chief (Section A.4.3). This disparity occurs due to the aggregation of managers that are not Yoruba, Hausa or Igbo into the "Other" category. Doing this leads to an overestimate of the proportion of firms with a similar ethnicity. In the estimations, the variable describing whether or not the firm manager shares the same ethnicity as the local government chief ("Similar-Ethnicity") is constructed using data on ethnicities without aggregation. Table A.74 shows the aggregated data for ease of explanation.

Constructed Indices Information on the corruption measures for each state are shown in Table A.75. Table A.76 shows the diversity measures for each state. Figure 22 plots information about 3 of the corruption measures: CI(2), CRI and CR. The data shows a relatively strong positive correlation between the three measures which helps with the argument that they are picking up similar things.

Measures Of Corruption Overall 28.7% of the 1,267 firms reported paying a bribe of some kind. CI(2) was constructed from businesses reporting the type of public official that they had paid bribes to. Firms were asked about a total of 11 types of public official. These options are not an exhaustive list of every type of person that a business might pay a bribe to, however, it does help to see the heterogeneity in bribe behaviour with different types of government officials; and to see the control rights that each type of official commands over the business. Also, the list of 11 variables used to construct CI(2) is similar to the set of variables used in previous literature [Hunt , 2004, 2006, 2007, 2008, Hunt & Laszlo , 2012]. 170 businesses (out of 1,267) reported that they had bribed at least one of the 11 types of officials. The maximum number of officials bribed was 9 (out of a possible 11). The mean average was 2.65 for those who reported bribing one of the 11 official-types; and 0.36 for the entire sample.

This study utilises the use of an official as a measure for the level of transactions in the economy. CRI is constructed by dividing the number of official types bribed by the number of official types used. Therefore, CRI can take on values between 0 and 1 inclusive. Values for CRI ranged from 0 to 0.81 with a mean of 0.31. The mean average for businesses who reported bribing one of the 11 official types is 0.42.

CR was generated by summing the amount of bribe collected in each region, this is a Naira figure and can be any non-negative value. The median regional total bribe revenue was ₦151,000 (mean=₦542,000). Excluding states with a total bribe revenue of more than ₦2,000,000 reduces these figures to ₦133,000 (mean=₦190,000).

Figure 22 shows a scatter plot of the CI(2) and CRI measures of corruption incidence and relative corruption incidence, respectively, for each region. The symbols are weighted by the CR measure of corruption revenue, so larger circles denote a larger amount received in total revenue within the state. The data shows a relatively strong positive relationship between CI(2) and CRI. The correlation coefficient between the two values is 0.76 which is statistically significant at the 1% level. Most regions lie within the 95% confidence regions with the exceptions of Rivers state (above) and Zamfara state (below). I.e. Rivers state has a higher corruption index than predicted by the data and Zamfara has a lower corruption index than predicted by the data.

Corruption And (Ethnic/Religious/Linguistic) (Polarisation/Fractionalisation): State Level Results Results from the regressions of corruption on diversity are shown in Tables A.77, A.78, A.79, A.80, A.81, and A.82.

Despite the fact that ethnicity data is only available for a sub-sample of firms, corruption data is available for all firms. The diversity measures in the aforementioned tables are constructed using individual level data from the CLEEN survey. Tables A.77,A.79 and A.81 use the full sample of firms to generate the corruption indices while Tables A.78, A.80 andA.82 use the sub-sample of 1,267 firms which have ethnicity data.

Table A.77 shows the results from regressing the four main corruption measures (CI(1); CI(2); CRI; CR) on the 3 unweighted measures of diversity (ethnic; religious; and linguistic fractionalisation). The measures of ethnic and religious fractionalisation are constructed in the traditional way whilst the measure of linguistic fractionalisation is created using the unweighted split personality method. Results are shown for the four corruption measures: 2 absolute corruption indices; the relative corruption index; and the log of corruption rents. The regressions also control for the share of firms that are in the manufacturing sector and the share of firms that engage in international trade. Observations are at the state level. Results show no statistically significant effect of any fractionalisation measure on any measure of corruption. All but one of the models are not statistically significant at the 10% level.

Similar results occur when using the subsample of 1,267 firms to generate the corruption indices. This is done in Table A.78, in these models none of the diversity variables return significant coefficients. One result that occurs quite often is the positive and statistically significant coefficient on the variable for the share of firms that engage in trade. This suggests that one factor influencing the level of corruption in an economy is international trade. More specifically, the factors which influence whether or not a firm engages in the direct importation of export of good and/or services are positively and significantly associated with level of corruption measured as the proportion of bribing firms; the respective number and proportion of transactions that attract bribes; and the total amount of bribes paid. The variable denoting trade has a positive association with the propensity to pay a bribe (CI(1)) and the amount of bribe paid (CR).

We see similar results when using polarisation as a measure of diversity (Table A.79). The majority of the models show no statistically significant effect of polarisation on corruption. Models V and VIII show negative effects of religious polarisation on the second corruption index and the relative corruption index that are significant at the 5% and 10% levels, respectively. These coefficients retain their sign and become more statistically significant when focusing on the sub-sample of 1,267 firms (Table A.80).

Tables A.81 and A.82 look at the effect of the different (weighted and unweughted) measures of linguistic fractionalisation on the different indices of corruption. The models of Table A.81 do not return many significant coefficients on the indices of linguistic diversity. Only 1 model (Table A.82, Model II) shows a positive effect of linguistic fractionalisation on corruption (Corruption Index 1). This result occurs when using the unweighted measure of linguistic fractionalisation that does not take into account the distances between languages. When weighting linguistic fractionalisation this effect disappears. Furthermore, all other models in this table show no significant effect of linguistic fractionalisation on corruption. The effect of trade on corruption amongst states in Nigeria is positive and statistically significant for all models in Tables A.81 and A.82.

Corruption And (Ethnic/Religious/Linguistic) (Polarisation/Fractionalisation): Local Government Level Results Tables A.83, A.84, A.85, A.86, A.87 and A.88 show results from the estimations run at the local government level. In general, these estimates return a greater number of statistically significant coefficients than those run at the state level. Moreover, many of the significant coefficients are in the models that use the relatively simpler measures of corruption - the corruption index; and the corruption rents. The corruption index (1) represents the proportion of bribing firms; the corruption rents represent the amount of money collected in bribes. These measures are used more often in the literature than CI (2) and CRI.

Table A.83, uses the full sample of firms to regress the different measures of corruption on diversity at the local government level. Model I shows a statistically significant relationship between ethnic fractionalisation (EF) & EF squared and corruption. The coefficients suggest a non-linear relationship between diversity and corruption.

$$\begin{aligned} CI &= \alpha_1 + \alpha_2 EF + \alpha_3 EF^2 + \mathbf{X}'\beta + \epsilon_i \\ \frac{\delta CI}{\delta EF} &= \alpha_2 + (2 * \alpha_3 EF) = 0.341 + (2 * (-0.361))EF \\ \frac{\delta CI}{\delta EF} &= 0.341 - 0.722EF \end{aligned} \tag{A.4.3.15}$$

In this sample, CI (1) ranges from 0 to 1 with an mean average value of 0.320 and a standard deviation of 0.354 (Table A.71). At this level of EF, an infinitesmally small increase in ethnic fractionalisation will lead to a 0.11 increase in the proportion of bribing firms (CI(1)). Therefore, at the average level of ethnic fractionalisation amongst local government areas in Nigeria, there seems to be evidence of a positive association between ethnic diversity and the occurrence of corruption.

Despite this, the model also suggests that this relationship reaches a peak where $EF = 0.472$, after this point an increase in the diversity of a local government area is associated with a fall in the incidence of corruption. This suggests that the capital of Abuja, with an EF value of 0.807 (Table A.76) will see a fall in this type of corruption if migrants from different parts of the country continue to flow into its borders. However, Lagos, with an EF of 0.382 will see an increase in the level of bribery by companies if it becomes more diverse than it is at the moment. This relationship is shown in Figure 23 which plots the fitted values from the estimation against EF. An inverse-U shaped relatiosnhip can be seen in the graph. Hence, the results of this study mirror those of Oguzhan [2008], who also finds an inverse U-shaped relatiosnhip between ethnic fractionalisation and corruption.

Model X of Table A.83 shows a similar result for the effect of EF on the amount of money collected in rents. In this case, the marginal effect of a small change in the level of fractionalisation is:

$$\begin{aligned}
Log(CR) &= \beta_1 + \beta_2 EF + \beta_3 EF^2 + \mathbf{X}'\gamma + \upsilon_i \\
\frac{\delta Log(CR)}{\delta EF} &= \beta_2 + (2 * \beta_3 EF) = 5.646 + (2 * (-5.3))EF \\
\frac{\delta Log(CR)}{\delta EF} &= 5.646 - 10.6EF
\end{aligned} \tag{A.4.3.16}$$

Thus, at the mean EF values of 0.320, a small change in the EF is associated with an increase in the log of bribe payments of 2.254. This is equivalent to an increase of ₦180,000 which is equal to £720. This function reaches a maximum at an EF level of 0.533, after which any increase in EF is associated with a reduction in the total amount of bribes paid. As before, Lagos finds itself on the positively sloped side of the curve while Abuja is on the negatively sloped side.

Similar results are found for the effect of religious diversity on the amount of money collected in corruption rents. In Model XI of Table A.83 both religious fractionalisation and its squared term have statistically significant coefficients. Starting at the mean value of religious fractionalisation (0.259) a 1 standard deviation (0.173) increase in the level of fractionalisation is associated with an increase in the total amount of Naira collected in bribes of log(0.382), which is equivalent to an increase of ₦1,500. Similar results are found when looking at the reduced sample (table A.84), however, the coefficients on religious fractionalisation and its square are insignificant for these estimations.

The results from tables A.85 and A.86, which look at the effect of the different measures of polarisation on the different measures of corruption, suggest that ethnic polarisation and linguistic polarisation work in opposite directions. Most of the coefficients on ethnic polarisation are positive whilst most of the coefficients on linguistic polarisation are negative. For religious polarisation, half of the coefficients are positive and half are negative, however, the only statistically significant coefficients are positive. These results suggest that higher ethnic and religious polarisation are associated with higher levels of corruption, in terms of the proportion of bribe paying firms and the total amount of Naira collected in bribes, while higher linguistic polarisation is associated with lower levels of corruption. The standard explanation for ELF and corruption can be used to explain the positive coefficients on ethnic and religious polarisation, respectively. An explanation for the negative coefficient on linguistic polarisation could be that, ignoring English, the lack of a common language might be a barrier to engaging in a (corrupt) transaction. In this setting the lack of a common language might drive distrust between both parties and might reduce the chance that a corrupt transaction takes place due to the possibility of being reported by the other party.

The coefficient on the share of firms in the local government area in the manufacturing sector is not statistically different from zero in the models that have a corruption index as a dependent variable. However, the coefficients on the share of manufacturing firms is positive and statistically significant in models X to XII of Table A.86 (the reduced sample). The coefficient on the "trade" variable is positive and statistically significant in models I to VI of both tables A.85 and A.86. Among all local government areas, the average proportion of firms that engaged in some form of international trade as part of ther operations was 5%. Therefore, the results from table A.86 indicate that, on average and ceteris paribus, a 10 percentage point higher share of firms engaging in international trade is associated with a 5 percentage point higher rate of bribery. The models seem best suited to explaining the models with corruption indices CI1 and CI2; and the corruption rents as dependent variables. The models seem least able to explain the relative corruption index. The models explaining the corruption index (CI1) and corruption rents are all statistically significant whilst the highest goodness of fit comes from the models explaining the second corruption index, CI2 (R-Squared=0.18).

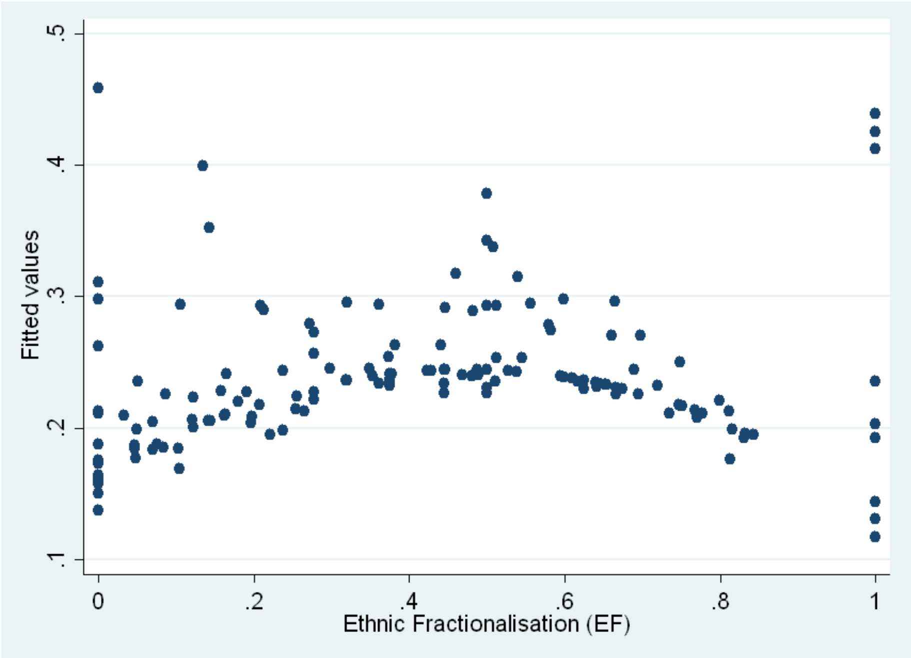
Tables A.87 and A.88 show the results from the estimations of corruption on the weighted and unweighted measures of linguistic fractionalisation. The coefficients on trade are positive and statistically significant in explaining the corruption indices (CI1 and CI2). The average number of languages spoken does not seem to be a significant explanatory variable when trying to explain the differences in the corruption levels across local government areas. Most of the coefficients for the average number of languages spoken are positive but none are statistically significant. The coefficients on the unweighted measure of linguistic fractionalisation (LF) and the weighted measure using the split personality method with $\delta = 0.5$ (LF_{05}) are not statistically significant in the models explaining the corruption indices (CI1 and CI2) and the relative corruption index. The coefficients on LF and LF_{05} are positive and statistically significant when entered into the models explaining the corruption rents. This suggests that, for a fixed number of languages, the more equally spread that the instances of the language are, the higher the total amount of money collected as bribe payments. Alternatively, other things being equal, the more languages spoken, the larger the amount of money collected in bribes.

The coefficients on the weighted measure of linguistic fractionalisation with $\delta = 0.05$ (LF_{005}) are negative in the models explaining the corruption indices and positive in the model explaining the amount of bribes collected. Therefore, linguistic polarisation is negatively associated with the total amount of bribes paid while linguistic fractionalisation has a positive effect. Despite this result, it might be useful to keep in mind that the result for linguistic polarisation is only significant at the 5% level when using the full sample of firms to measure corruption. When using the reduced sample the coefficient on linguistic polarisation is not significant. Nevertheless, looking at the data aids in seeing a potential reason for this result. This is done in Figure 24.

The y-axis displays the index number while the x-axis displays the number of groups. "Fractionalisation" and "Polariation" show the indices consistent with groups of equal size. I.e. "Fractionalisation" shows how the unweighted index of fractionalisation changes as the number of group sizes increases, keeping the share of each group within the population equal between groups. The same goes for "Polarisation". The difference between this theoretical trend and the observed values for linguistic fractionalisation (LFU) and polarisation (LP) indicates the effect of differing group sizes. So the spread of instances of languages is not equal among the population. In addition to this, under the theoretical indices, fractionalisation and polarisation move closer together for when there are more than 2 groups. This is not the case with the observed data, which shows a divergence of LFU and LP as the number of groups increases. A similar story emerges when the weighted measures of fractionalisation are plotted as well. A potential reason for the opposite effects of linguistic fractionalisation and polarisation on the amount of money collected in bribes is that the two variables tend to be moving in opposite directions in Figure 24.

Another potentially interesting result it that the negative effect of linguistic fractionalisation on the corruption indices is only significant when using the weighted measure of linguistic fractionalisation with $\delta = 0.05$. This study now examines each of these results in turn. Relatively low levels of δ seperate the languages that have no (or very few) branches in common from the rest of the languages, higher values of δ give a greater weight to smaller difference, while larger differences play a relatively less important role. In the limit, as $\delta \to \infty$, the measure of fractionalisation becomes one of polarisation and the most minute difference between languages is treated as complete seperateness and is not distinguishable from greater linguistic differences. The results from Table A.88 suggest that it is the deeper linguistic cleavages that matter when trying to explain the effect of diversity on the level of bribery rather than smaller differences in language. The coefficient for the effects of weighted linguistic diversity ($\delta = 0.05$) on the corruption indices are the opposite sign of the coefficient for the effect of linguistic diversity on the amount of bribe rents. This is consistent with previous literature which finds an inverse relationship between the propensity to bribe and the amount of bribe paid [Shleifer & Vishny , 1993]. In this case, the factors af-

Figure 23: A Plot Of Fitted Values For The Corruption Index (1) Against Ethnic Fractionalisation For Model I Of LG Estimations (Full Sample)

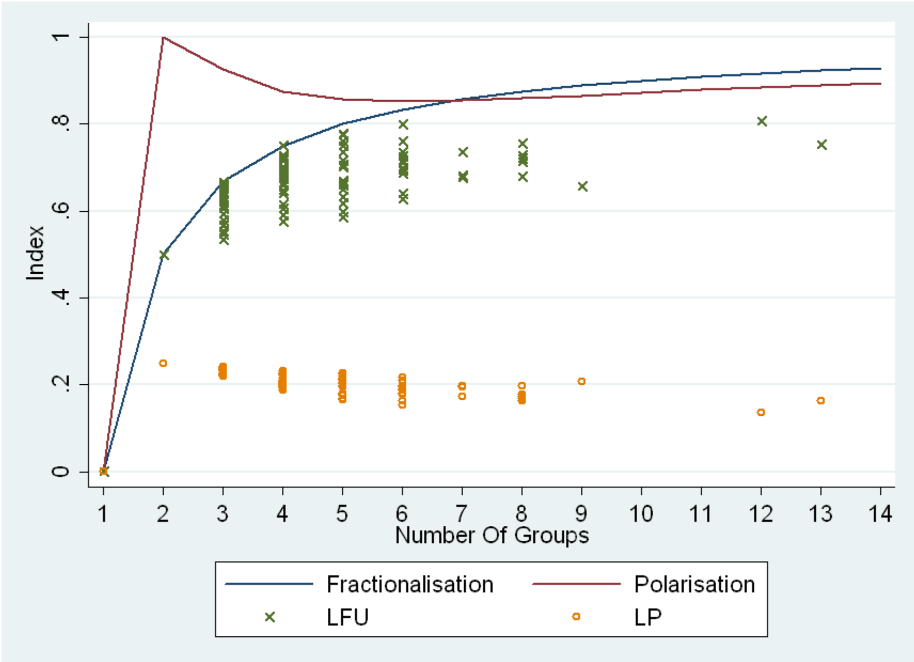


fecting the aggregate propensity to bribe have an opposite effect in explaining the size of bribe payment.

Conclusion This study has added to the body of knowledge by investigating the effect of ethnic; religious; and linguistic diversity on corruption. The analysis focused on businesses in the Nigerian economy. One of the main contributions of this study was to seperate ethnic diversity from linguistic diversity, this was achieved by finding out the respective ethnic group of; and languages spoken by, business managers. These were used along with information on the ethnic and religious mix of the different regions in the country to construct state level indices of polarisation and fractionalisation based on ethnicity; religion; and language. This study also accounted for the similarity between languages by constructing weighted indices of linguistic fractionalisation. This added to the literature by providing a measure of how related the languages were. Also, by using information about agents (individuals and firms) and not regions; this study was able to overcome the measurement error that occurs when assigning an ethnic/religious identity to all of the individuals located in a region.

Four separate measures of corruption were constructed which were based on those used or described by previous literature. These were: the proportion of

Figure 24: The Relationship Between Fractionalisation And Polarisation: Theoretical And Empirical Results



Y-axis: index value. X-axis: number of groups in population. The blue (red) line represents the relationship between the index of fractionalisation (polarisation) and the number of groups within the population, holding the share of each group within the population equal among groups. The green X' (orange circles) show the actual data on fractionalisation (polarisation), where the groups shares are can vary.

bribing firms in a region; the average number of bribe types paid per firm in a region; the average number of bribe types paid as a fraction of total possible bribe types; and the (log of the) total amount of bribes collected, in Naira, in a region. These were defined as: corruption index (1); corruption index (2); relative corruption index; and corruption rents, respectively.

Results show no robust relationship between the measures of diversity and the measures of corruption. There was some evidence for a negative effect of religious and linguistic polarisation on corruption. There was also some weak evidence for a positive effect of unweighted linguistic fractionalisation on corruption. This result suggests that the positive effect of Ethnolinguistic Fractionalisation on corruption that is found or assumed in the literature might be a result of: conflating two seperate things, ethnicity and linguism; and treating all groups as being equally distant from one another. When weighting linguistic groups by their distance to each other, using a language tree, no significant relationship shows up between fractionalisation and corruption. These results remained when using measures of ethnic fractionalisation that encorporated racial distance. Future work can use other methods to calculate ethno-linguistic diversity, e.g. random-speaker method (weighted and unweighted); random speaker-hearer method; and index of communication. Comparing results from the different measures will aid in determining the best one to use when looking at the effect of diversity on corruption.

Table A.75: Corruption Measures, By Region

Region	CI(1)	CI(2)	CRI	CR	LCR
Abia	0.326	1.087	0.615	199	5.291
Adamawa	0.235	0.294	0.333	883	6.784
Akwa ibom	0.323	0.710	0.808	78	4.361
Anambra	0.317	0.878	0.459	174	5.159
Bauchi	0.250	0.350	0.252	151	5.020
Bayelsa	0.100	0.450	0.500	20	2.996
Benue	0.286	0.143	0.094	5	1.674
Borno	0.300	0.400	0.311	125	4.828
Cross River	0.083	0.167	0.333	2	0.693
Delta	0.242	0.394	0.407	81	4.388
Ebonyi	0.313	0.625	0.503	202	5.308
Edo	0.148	0.185	0.126	387	5.960
Ekiti	0.038	0.077	0.100	14	2.645
Enugu	0.290	0.355	0.567	381	5.942
Gombe	0.080	0.120	0.156	3094	8.037
Imo	0.137	0.275	0.262	55	4.002
Jigawa	0.000	0.000	.	0	.
Kaduna	0.154	0.269	0.211	260	5.560
Kano	0.154	0.000	0.000	109	4.689
Katsina	0.333	0.444	0.579	3299	8.102
Kebbi	0.167	0.917	0.667	55	4.007
Kogi	0.429	0.048	0.077	279	5.631
Kwara	0.346	0.000	0.000	447	6.101
Lagos	0.582	0.432	0.337	7180	8.879
Nassarawa	0.200	0.040	0.014	441	6.088
Niger	0.389	0.056	0.250	142	4.954
Ogun	0.029	0.235	0.143	1	0.693
Ondo	0.037	0.000	0.000	158	5.064
Osun	0.051	0.051	0.200	34	3.518
Oyo	0.175	0.125	0.317	101	4.616
Plateau	0.326	0.279	0.114	725	6.586
Rivers	0.333	1.381	0.587	495	6.205
Sokoto	0.300	0.800	0.389	251	5.525
Taraba	0.111	0.000	0.000	22	3.110
Yobe	0.000	0.000	.	0	.
Zamfara	0.118	0.176	0.667	23	3.135
FCT (Abuja)	0.118	0.088	0.100	177	5.178

CI(1): The proportion of bribing firms in the region. CI(2): The regional average number of different purposes for which a bribe was paid (potential max=11). CRI: The regional average of: The number of different purposes for which a bribe was paid as a fraction of the number of different purposes for which a bribe could have been paid (i.e. the number of bribe types divided by the number of different types of public officials met with). CR: The total amount of bribes paid by firms within the region (in '000 Naira). LCR: Log of: the total amount of bribes paid by firms within the region.

Table A.76: Diversity Measures, By Region

Region	Ethnic Polarisation	Religious Polarisation	Linguistic Polarisation	Ethnic Fractionalisation (Unweighted)	Ethnic Fractionalisation ($\delta = .5$)	Ethnic Fractionalisation ($\delta = .05$)	Religious Fractionalisation	Linguistic Fractionalisation (Unweighted)	Linguistic Fractionalisation ($\delta = .5$)	Linguistic Fractionalisation ($\delta = .05$)
Abia	0.035	0.023	0.221	0.074	0.014	.002	0.046	0.607	0.559	0.524
Adamawa	0.158	0.220	0.196	0.477	0.093	0.010	0.457	0.693	0.670	0.651
Akwa Ibom	0.187	0.010	0.192	0.742	0.136	0.015	0.021	0.708	0.565	0.504
Anambra	0.033	0.016	0.227	0.070	0.012	0.001	0.032	0.591	0.548	0.524
Bauchi	0.134	0.152	0.210	0.361	0.077	0.008	0.308	0.655	0.646	0.636
Bayelsa	0.162	0.030	0.184	0.461	0.051	0.006	0.061	0.706	0.603	0.505
Benue	0.183	0.037	0.185	0.605	0.119	0.013	0.075	0.709	0.638	0.587
Borno	0.168	0.178	0.201	0.742	0.199	0.068	0.363	0.704	0.684	0.666
Cross River	0.175	0.009	0.152	0.633	0.103	0.011	0.018	0.785	0.576	0.482
Delta	0.170	0.026	0.182	0.440	0.052	0.006	0.053	0.726	0.574	0.494
Ebonyi	0.006	0.109	0.250	0.012	0.003	0.000	0.237	0.500	0.500	0.500
Edo	0.135	0.123	0.180	0.821	0.055	0.006	0.265	0.720	0.580	0.510
Ekiti	0.066	0.078	0.242	0.142	0.014	0.001	0.158	0.535	0.513	0.501
Enugu	0.031	0.094	0.213	0.065	0.012	0.001	0.193	0.630	0.566	0.530
Gombe	0.186	0.219	0.207	0.541	0.119	0.013	0.453	0.656	0.645	0.635
Imo	0.017	0.021	0.211	0.035	0.005	0.001	0.041	0.642	0.576	0.523
Jigawa	0.088	0.059	0.238	0.203	0.045	0.005	0.120	0.548	0.548	0.547
Kaduna	0.154	0.239	0.193	0.454	0.096	0.011	0.484	0.705	0.677	0.662
Kano	0.121	0.133	0.231	0.296	0.064	0.007	0.272	0.589	0.589	0.589
Katsina	0.040	0.034	0.218	0.084	0.019	0.002	0.069	0.619	0.612	0.607
Kebbi	0.090	0.097	0.195	0.208	0.042	0.005	0.200	0.697	0.677	0.659
Kogi	0.174	0.249	0.139	0.727	0.084	0.009	0.503	0.800	0.643	0.547
Kwara	0.051	0.248	0.221	0.108	0.014	0.002	0.505	0.609	0.577	0.552
Lagos	0.155	0.216	0.206	0.382	0.045	0.005	0.444	0.661	0.572	0.521
Nassarawa	0.145	0.221	0.173	0.798	0.151	0.017	0.462	0.731	0.695	0.669
Niger	0.163	0.208	0.204	0.617	0.127	0.014	0.430	0.675	0.658	0.649
Ogun	0.056	0.243	0.247	0.121	0.013	0.001	0.507	0.514	0.506	0.501
Ondo	0.087	0.077	0.236	0.187	0.018	0.002	0.158	0.551	0.520	0.503
Osun	0.033	0.249	0.241	0.068	0.009	0.001	0.498	0.536	0.512	0.501
Oyo	0.109	0.229	0.231	0.244	0.028	0.003	0.469	0.574	0.528	0.503
Plateau	0.155	0.049	0.164	0.756	0.143	0.016	0.099	0.749	0.698	0.661
Rivers	0.172	0.042	0.178	0.738	0.098	0.011	0.088	0.711	0.603	0.511
Sokoto	0.134	0.091	0.240	0.325	0.070	0.008	0.189	0.544	0.544	0.544
Taraba	0.158	0.189	0.202	0.771	0.143	0.016	0.378	0.684	0.670	0.656
Yobe	0.153	0.161	0.225	0.715	0.135	0.015	0.333	0.599	0.595	0.593
Zamfara	0.110	0.101	0.215	0.246	0.054	0.006	0.216	0.631	0.619	0.604
FCT (Abuja)	0.136	0.186	0.178	0.807	0.120	0.013	0.373	0.737	0.676	0.642

Table A.77: State Level Estimations Of Corruption On Fractionalisation - Full Sample

Dependent Variable:	Corruption Index			Corruption Index (2)			Relative Corruption Index			Log Of Corruption Rents		
	1	2	3	4	5	6	7	8	9	10	11	12
Ethnic Fractionalisation (EF)	-0.042			-0.379			-0.272			1.833		
	(0.283)			(0.479)			(0.363)			(3.440)		
EF Squared	0.047			0.216			-0.071			-1.389		
	(0.304)			(0.568)			(0.424)			(3.694)		
Religious Fractionalisation (RF)		-0.208			-0.640			-0.324			3.764	
		(0.527)			(0.899)			(0.895)			(4.345)	
RF Squared		0.746			0.320			0.012			-2.539	
		(0.968)			(1.579)			(1.607)			(7.019)	
Linguistic Fractionalisation (Unweighted) (LFU)			2.514			2.882			8.808*			43.254
			(5.604)			(4.995)			(5.145)			(31.400)
LFU Squared			-1.830			-2.638			-7.428*			-32.158
			(4.455)			(3.860)			(3.994)			(24.829)
Manufacture (Share)	-0.341*	-0.399*	-0.297	-1.274***	-0.932**	-1.356***	-0.990***	-0.653**	-1.067***	-3.511**	-5.372***	-3.313*
	(0.185)	(0.197)	(0.201)	(0.359)	(0.421)	(0.388)	(0.223)	(0.288)	(0.204)	(1.614)	(1.735)	(1.695)
Trade (share)	1.212*	1.467**	1.108	3.170***	2.327**	3.371***	1.777***	1.211	2.095***	15.939***	20.960***	15.310***
	(0.634)	(0.660)	(0.656)	(0.974)	(1.054)	(0.917)	(0.649)	(0.857)	(0.615)	(3.966)	(4.150)	(3.911)
Constant	0.179***	0.149**	-0.678	0.375***	0.406***	-0.470	0.437***	0.376***	-2.241	4.667***	4.317***	-9.257
	(0.055)	(0.060)	(1.739)	(0.085)	(0.103)	(1.570)	(0.068)	(0.094)	(1.621)	(0.548)	(0.632)	(9.705)
R-squared	0.153	0.234	0.171	0.323	0.364	0.301	0.380	0.246	0.304	0.302	0.378	0.317
Observations	37	37	37	37	37	37	37	37	37	35	35	35
F-Stat	1.24	1.61	1.46	7.76	3.42	4.56	9.43	3.50	7.90	5.56	7.38	7.60
P-value	0.313	0.195	0.237	0.000	0.019	0.005	0.000	0.018	0.000	0.002	0.000	0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$

Models are estimated using an OLS regression. Huber-White Standard Errors In Parenthesis. Number of Observations for Models 1-9: 37. These 37 observations correspond to the 37 geo-political states in Nigeria. The 37 states are generated using data on all 2,110 firms in the sample. Models 10-12 contain 35 observations due to the dependent variable [log(0)] being undefined for Jigawa State and Yobe State.

Table A.78: State Level Estimations Of Corruption On Fractionalisation - Reduced Sample

Dependent Variable:	Corruption Index			Corruption Index (2)			Relative Corruption Index			Log Of Corruption Rents		
	1	2	3	4	5	6	7	8	9	10	11	12
Ethnic Fractionalisation (EF)	0.016			-0.376			0.068			-0.524		
	(0.297)			(0.682)			(0.457)			(4.418)		
EF Squared	0.080			0.329			-0.287			2.058		
	(0.338)			(0.797)			(0.552)			(4.665)		
Religious Fractionalisation (RF)		-0.601			-1.053			-0.626			7.462	
		(0.486)			(1.148)			(0.909)			(8.007)	
RF Squared		1.287			0.605			0.162			-8.243	
		(0.995)			(1.993)			(1.628)			(15.776)	
Linguistic Fractionalisation (Unweighted) (LFU)			1.661			5.743			8.911			121.425
			(5.339)			(6.545)			(6.155)			(79.924)
LFU Squared			-0.895			-4.530			-6.986			-91.261
			(4.238)			(5.052)			(4.800)			(62.397)
Manufacture (Share)	-0.063	-0.028	-0.022	-0.789**	-0.634*	-0.822**	-0.625**	-0.423	-0.605**	1.447	-0.140	1.062
	(0.152)	(0.153)	(0.150)	(0.364)	(0.364)	(0.370)	(0.260)	(0.252)	(0.232)	(2.051)	(2.016)	(2.051)
Trade (share)	1.483***	1.503***	1.400***	3.193***	2.737***	3.196***	1.507*	1.221	1.552*	18.370**	19.318***	15.544***
	(0.536)	(0.502)	(0.506)	(1.024)	(0.921)	(1.030)	(0.811)	(0.766)	(0.840)	(6.837)	(5.146)	(5.606)
Constant	0.131*	0.183***	-0.541	0.365***	0.504***	-1.497	0.351***	0.445***	-2.493	3.552***	2.851***	-35.833
	(0.065)	(0.050)	(1.654)	(0.127)	(0.143)	(2.061)	(0.091)	(0.088)	(1.929)	(0.979)	(0.844)	(25.390)
R-squared	0.248	0.267	0.306	0.277	0.401	0.275	0.232	0.349	0.223	0.192	0.238	0.275
Observations	37	37	37	37	37	37	35	35	35	35	35	35
F-Stat	2.45	2.56	3.39	3.83	4.37	3.73	3.21	4.10	3.27	2.04	4.31	2.29
P-value	0.066	0.057	0.020	0.012	0.006	0.013	0.026	0.009	0.025	0.114	0.007	0.083

$*p < 0.10, **p < 0.05, ***p < 0.01$

Models are estimated using an OLS regression. Huber-White Standard Errors In Parenthesis. Number of Observations for Models 1-6: 37. These 37 observations correspond to the 37 geo-political states in Nigeria. The 37 states are generated using data on 1,267 firms for which ethnicity, religion, and language data were available. Models 7-9: 35 Observations due to the dependent variable being undefined for Jigawa and Yobe (division by zero). Models 10-12: 35 observations due to the dependent variable [log(0)] being undefined for Jigawa State and Yobe State.

Table A.79: State Level Estimations Of Corruption On Polarisation - Full Sample

Dependent Variable:	Corruption Index			Corruption Index (2)			Relative Corruption Index			Log Of Corruption Rents		
	1	2	3	4	5	6	7	8	9	10	11	12
Ethnic Polarisation	0.025			-0.861			-1.267**			3.506		
	(0.345)			(0.613)			(0.521)			(3.657)		
Religious Polarisation		0.397			-0.967**			-0.660*			4.823*	
		(0.297)			(0.370)			(0.353)			(2.565)	
Linguistic Polarisation			-0.389			1.411			2.448***			-3.899
			(0.828)			(0.941)			(0.792)			(4.930)
Manufacture (Share)	-0.341*	-0.443**	-0.299	-1.254***	-0.947**	-1.350***	-0.910***	-0.651**	-1.093***	-3.495**	-5.226***	-3.370**
	(0.175)	(0.183)	(0.196)	(0.386)	(0.394)	(0.367)	(0.238)	(0.277)	(0.198)	(1.515)	(1.581)	(1.643)
Trade (share)	1.202*	1.521**	1.130*	3.226***	2.337**	3.381***	1.902***	1.201	2.203***	15.811***	20.800***	15.748***
	(0.630)	(0.640)	(0.665)	(0.930)	(1.021)	(0.885)	(0.696)	(0.843)	(0.614)	(3.878)	(4.005)	(4.230)
Constant	0.170***	0.120***	0.252	0.368***	0.394***	-0.019	0.440***	0.377***	-0.204	4.673***	4.426***	5.868***
	(0.051)	(0.044)	(0.172)	(0.080)	(0.081)	(0.200)	(0.066)	(0.064)	(0.164)	(0.497)	(0.447)	(1.017)
R-squared	0.153	0.218	0.159	0.312	0.366	0.295	0.306	0.248	0.272	0.302	0.373	0.281
Observations	37	37	37	37	37	37	37	37	37	35	35	35
F-Stat	1.43	2.11	1.35	6.76	4.77	7.19	9.19	4.79	14.74	6.81	9.39	6.23
P-value	0.250	0.117	0.276	0.001	0.007	0.001	0.000	0.007	0.000	0.001	0.000	0.002

$*p < 0.10, **p < 0.05, ***p < 0.01$

. Models are estimated using an OLS regression. Huber-White Standard Errors In Parenthesis. Number of Observations models 1-9: 37. These 37 observations correspond to the 37 geo-political states in Nigeria. The 37 states are generated using data on all 2,110 firms in the dataset. Models 10-12: 35 observations due to the dependent variable [log(0)] being undefined for Jigawa State and Yobe State.

Table A.80: State Level Estimations Of Corruption On Polarisation - Reduced Sample

Dependent Variable:	Corruption Index			Corruption Index (2)			Relative Corruption Index			Log Of Corruption Rents		
	1	2	3	4	5	6	7	8	9	10	11	12
Ethnic Polarisation	0.400			-0.399			-0.422			3.818		
	(0.365)			(0.929)			(0.629)			(5.999)		
Religious Polarisation		0.177			-1.509***			-1.121***			6.217	
		(0.283)			(0.469)			(0.310)			(4.003)	
Linguistic Polarisation			-1.446*			0.310			0.519			-10.274
			(0.774)			(1.560)			(1.194)			(13.735)
Manufacture (Share)	-0.080	-0.100	-0.009	-0.799**	-0.666*	-0.810**	-0.575**	-0.429*	-0.592**	1.104	0.281	1.460
	(0.151)	(0.154)	(0.149)	(0.371)	(0.338)	(0.348)	(0.235)	(0.238)	(0.234)	(2.073)	(1.802)	(1.945)
Trade (share)	1.448***	1.452***	1.429***	3.230***	2.704***	3.286***	1.646*	1.203	1.686**	17.552**	19.758***	17.522**
	(0.509)	(0.512)	(0.509)	(0.983)	(0.895)	(1.016)	(0.812)	(0.737)	(0.826)	(6.593)	(5.362)	(6.468)
Constant	0.114*	0.141***	0.450***	0.338***	0.486***	0.226	0.347***	0.442***	0.192	3.463***	3.125***	5.969**
	(0.060)	(0.046)	(0.161)	(0.117)	(0.121)	(0.342)	(0.087)	(0.073)	(0.241)	(0.948)	(0.651)	(2.776)
R-squared	0.247	0.232	0.300	0.272	0.403	0.269	0.197	0.353	0.189	0.174	0.231	0.181
Observations	37	37	37	37	37	37	35	35	35	35	35	35
F-Stat	3.05	2.90	3.89	4.77	6.00	5.17	3.71	5.62	3.49	2.46	5.62	2.47
P-value	0.042	0.050	0.017	0.007	0.002	0.005	0.022	0.003	0.027	0.081	0.003	0.081

$*p < 0.10, **p < 0.05, ***p < 0.01$

. Models are estimated using an OLS regression. Huber-White Standard Errors In Parenthesis. Number of Observations for models 1-6: 37. These 37 observations correspond to the 37 geo-political states in Nigeria. The 37 states are generated using data on 1,267 firms for which ethnicity, religion, and language data were available. Models 7-9: 35 Observations due to the dependent variable being undefined for Jigawa and Yobe (division by zero). Models 10-12: 35 observations due to the dependent variable [log(0)] being undefined for Jigawa State and Yobe State.

Table A.81: State Level Estimations Of Corruption On Linguistic Fractionalisation - Full Sample

Dependent Variable:	Corruption Index				Corruption Index (2)				Relative Corruption Index				Log Of Corruption Rents			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Avg. No. Of Languages Spoken	0.011				-0.151				-0.114				0.663			
	(0.082)				(0.191)				(0.143)				(0.520)			
Linguistic Fractionalisation (Unweighted)(LFU)		0.163				-0.507				-0.738**				1.910		
		(0.254)				(0.326)				(0.296)				(1.566)		
Linguistic Fractionalisation $(\delta = .5)(LF_{05})$			0.151				-0.846				-0.860*				5.031***	
			(0.256)				(0.574)				(0.479)				(1.816)	
Linguistic Fractionalisation $(\delta = .05)(LF_{005})$				0.080				-0.853				-0.608				5.667**
				(0.257)				(0.579)				(0.477)				(2.166)
Manufacture (Share)	-0.336*	-0.292	-0.330*	-0.347*	-1.281***	-1.349***	-1.263***	-1.150***	-0.885***	-1.048***	-0.891***	-0.788***	-3.440**	-3.236*	-3.424**	-4.080***
	(0.178)	(0.191)	(0.179)	(0.176)	(0.439)	(0.385)	(0.401)	(0.383)	(0.239)	(0.209)	(0.262)	(0.263)	(1.571)	(1.609)	(1.487)	(1.453)
Trade (share)	1.199*	1.116*	1.190*	1.216*	3.193***	3.382***	3.190***	2.986***	1.792**	2.128***	1.811**	1.641*	16.182***	15.521***	16.099***	17.244***
	(0.644)	(0.654)	(0.643)	(0.633)	(0.986)	(0.928)	(0.966)	(0.951)	(0.765)	(0.669)	(0.782)	(0.810)	(4.134)	(4.089)	(3.862)	(3.998)
Constant	0.147	0.065	0.082	0.128	0.623	0.601***	0.777**	0.750**	0.558*	0.778***	0.809***	0.634**	3.509**	3.815***	2.036*	1.867
	(0.193)	(0.168)	(0.152)	(0.145)	(0.461)	(0.216)	(0.346)	(0.334)	(0.327)	(0.194)	(0.288)	(0.269)	(1.281)	(1.073)	(1.083)	(1.231)
R-squared	0.153	0.162	0.158	0.154	0.292	0.297	0.314	0.323	0.195	0.250	0.244	0.219	0.289	0.288	0.337	0.371
Observations	37	37	37	37	37	37	37	37	37	37	37	37	35	35	35	35
F-Stat	1.36	1.40	1.69	1.51	3.61	5.52	4.26	3.62	6.37	12.73	5.49	4.53	5.84	6.52	10.22	8.86
P-value	0.274	0.261	0.189	0.231	0.023	0.003	0.012	0.023	0.002	0.000	0.004	0.009	0.003	0.002	0.000	0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$

. Models are estimated using an OLS regression. Huber-White Standard Errors In Parenthesis. Number of Observations for models 1-12: 37. These 37 observations correspond to the 37 geo-political states in Nigeria. The 37 states are generated using data on all 2,110 firms in the dataset. Models 13-16: 35 observations due to the dependent variable [log(0)] being undefined for Jigawa State and Yobe State.

Table A.82: State Level Estimations Of Corruption On Linguistic Fractionalisation - Reduced Sample

Dependent Variable:	Corruption Index				Corruption Index (2)				Relative Corruption Index				Log Of Corruption Rents			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Avg. No. Of Languages Spoken	0.111				-0.014				0.079				0.467			
	(0.095)				(0.225)				(0.154)				(1.512)			
Linguistic Fractionalisation (Unweighted) (LFU)		0.511*				-0.077				-0.058				4.166		
		(0.255)				(0.515)				(0.411)				(4.730)		
Linguistic Fractionalisation ($\delta = .5$) (LF_{05})			0.341				-0.694				-0.571				8.013	
			(0.328)				(0.757)				(0.574)				(4.803)	
Linguistic Fractionalisation ($\delta = .05$) (LF_{005})				0.025				-1.081				-0.731				7.005
				(0.325)				(0.752)				(0.533)				(4.334)
Manufacture (Share)	-0.063	-0.018	-0.088	-0.088	-0.796*	-0.804**	-0.789**	-0.697*	-0.568**	-0.578**	-0.556**	-0.502**	1.106	1.421	0.827	0.396
	(0.145)	(0.145)	(0.151)	(0.165)	(0.392)	(0.359)	(0.377)	(0.353)	(0.220)	(0.230)	(0.245)	(0.246)	(2.014)	(1.914)	(1.872)	(2.023)
Trade (share)	1.416**	1.418***	1.356**	1.378**	3.292***	3.291***	3.348***	3.457***	1.757**	1.706**	1.737**	1.820**	17.155**	17.557**	16.603**	15.910**
	(0.525)	(0.503)	(0.545)	(0.543)	(1.023)	(1.016)	(0.974)	(1.009)	(0.849)	(0.836)	(0.766)	(0.758)	(6.655)	(6.446)	(6.534)	(6.707)
Constant	-0.098	-0.179	-0.039	0.150	0.320	0.339	0.700	0.881*	0.107	0.333	0.634*	0.695**	2.837	1.146	-0.826	0.100
	(0.221)	(0.169)	(0.185)	(0.171)	(0.564)	(0.336)	(0.460)	(0.444)	(0.363)	(0.278)	(0.352)	(0.304)	(3.675)	(3.245)	(3.027)	(2.526)
R-squared	0.252	0.304	0.242	0.220	0.268	0.268	0.283	0.309	0.192	0.187	0.209	0.231	0.164	0.188	0.222	0.216
Observations	37	37	37	37	37	37	37	37	35	35	35	35	35	35	35	35
F-Stat	3.24	4.23	3.14	2.72	4.40	4.74	5.02	4.78	2.88	3.19	3.60	3.77	2.31	2.49	2.37	2.43
p-value	0.035	0.012	0.038	0.060	0.010	0.007	0.006	0.007	0.052	0.037	0.024	0.020	0.096	0.079	0.090	0.084

$*p < 0.10, **p < 0.05, ***p < 0.01$

.Models are estimated using an OLS regression. Huber-White Standard Errors In Parenthesis. Number of Observations for models 1-8: 37. These 37 observations correspond to the 37 geo-political states in Nigeria. The 37 states are generated using data on 1,267 firms for which ethnicity, religion, and language data were available. Models 9-12: 35 Observations due to the dependent variable being undefined for Jigawa and Yobe (division by zero). Models 13-16: 35 observations due to the dependent variable [log(0)] being undefined for Jigawa State and Yobe State.

Table A.83: Local Government Level Estimations Of Corruption On Fractionalisation - Full Sample

Dependent Variable:	Corruption Index			Corruption Index (2)			Relative Corruption Index			Log Of Corruption Rents		
	1	2	3	4	5	6	7	8	9	10	11	12
Ethnic Fractionalisation (EF)	0.341**			0.320			0.006			5.646***		
	(0.148)			(0.274)			(0.221)			(1.278)		
EF Squared	-0.361**			-0.455*			-0.173			-5.300***		
	(0.157)			(0.257)			(0.223)			(1.291)		
Religious Fractionalisation (RF)		0.208			-0.032			-0.109			4.636***	
		(0.146)			(0.267)			(0.210)			(1.358)	
RF Squared		-0.245			-0.151			-0.061			-4.684***	
		(0.150)			(0.244)			(0.210)			(1.306)	
Linguistic Fractionalisation (LF)			-1.191***			0.170			0.100			-2.774
			(0.360)			(0.553)			(0.436)			(3.184)
LF Squared			1.261***			-0.031			-0.193			7.098*
			(0.388)			(0.623)			(0.570)			(3.604)
Manufacture (Share)	-0.027	-0.030	-0.014	0.051	0.066	0.094	-0.066	-0.039	-0.067	0.839	0.332	0.771
	(0.063)	(0.064)	(0.067)	(0.131)	(0.134)	(0.166)	(0.081)	(0.081)	(0.097)	(0.618)	(0.657)	(0.788)
Trade (share)	0.295**	0.300**	0.376**	0.840**	0.850**	0.742**	-0.016	-0.010	-0.072	1.731	1.840	3.163*
	(0.148)	(0.148)	(0.177)	(0.393)	(0.395)	(0.368)	(0.126)	(0.126)	(0.161)	(1.154)	(1.247)	(1.865)
Constant	0.164***	0.180***	0.443***	0.186***	0.215***	0.115	0.313***	0.307***	0.304***	2.404***	2.742***	2.057***
	(0.025)	(0.024)	(0.096)	(0.055)	(0.051)	(0.161)	(0.049)	(0.042)	(0.090)	(0.236)	(0.225)	(0.786)
R-squared	0.041	0.034	0.048	0.061	0.063	0.032	0.027	0.026	0.004	0.155	0.109	0.134
Observations	350	350	292	350	350	292	200	200	180	152	152	140
F-Stat	2.61	1.84	3.63	5.67	5.90	1.46	1.84	1.64	0.22	7.35	4.93	4.91
P-value	0.035	0.121	0.007	0.000	0.000	0.214	0.122	0.166	0.929	0.000	0.001	0.001

$*p < 0.10, **p < 0.05, ***p < 0.01$

Models are estimated using an OLS regression. Huber-White Standard Errors In Parenthesis. Number of Observations: 351. These 351 observations correspond to the 351 local government areas in the sample. The 351 local government areas are formed using data on all 2,110 firms in the sample. The models in this table correspond to the Models in Table A.77. Models 7 & 9: 150 Observations dropped due to division by zero. Models 10-12: 198 Oservations dropped due to log(0) being undefined.

Table A.84: Local Government Level Estimations Of Corruption On Fractionalisation - Reduced Sample

Dependent Variable:	Corruption Index			Corruption Index (2)			Relative Corruption Index			Log Of Corruption Rents		
	1	2	3	4	5	6	7	8	9	10	11	12
Ethnic Fractionalisation (EF)	0.311			0.230			0.406			4.710**		
	(0.212)			(0.464)			(0.320)			(2.197)		
EF Squared	-0.302			-0.329			-0.724*			-4.125		
	(0.291)			(0.634)			(0.412)			(2.937)		
Religious Fractionalisation (RF)		0.110			0.840			0.591			4.967	
		(0.324)			(0.747)			(0.526)			(3.361)	
RF Squared		0.054			-2.234			-1.540			-4.891	
		(0.656)			(1.456)			(1.017)			(6.506)	
Linguistic Fractionalisation (LF)			-2.891***			0.572			0.896*			-0.121
			(0.375)			(0.475)			(0.488)			(3.774)
LF Squared			2.697***			-0.375			-0.801			4.373
			(0.409)			(0.591)			(0.636)			(4.408)
Manufacture (Share)	-0.065	-0.072	-0.052	0.005	0.026	0.007	-0.100	-0.056	-0.088	2.456**	2.107*	2.137*
	(0.064)	(0.064)	(0.062)	(0.167)	(0.169)	(0.164)	(0.109)	(0.111)	(0.110)	(1.133)	(1.167)	(1.285)
Trade (share)	0.494***	0.495***	0.517***	2.178**	2.172**	2.186**	0.096	0.106	0.111	0.821	0.793	1.063
	(0.161)	(0.162)	(0.162)	(0.975)	(0.975)	(0.974)	(0.159)	(0.159)	(0.162)	(1.212)	(1.182)	(1.223)
Constant	0.165***	0.178***	0.936***	0.180***	0.198***	-0.013	0.304***	0.305***	0.061	2.324***	2.524***	1.453
	(0.028)	(0.026)	(0.097)	(0.056)	(0.050)	(0.149)	(0.055)	(0.048)	(0.098)	(0.305)	(0.281)	(0.903)
R-squared	0.077	0.072	0.097	0.175	0.180	0.175	0.025	0.024	0.011	0.135	0.126	0.112
Observations	292	292	292	292	292	292	160	160	160	119	119	119
F-Stat	3.83	3.30	17.64	1.66	2.25	3.32	1.67	1.42	1.86	5.11	5.16	3.93
P-value	0.005	0.012	0.000	0.160	0.064	0.011	0.160	0.229	0.121	0.001	0.001	0.005

$*p < 0.10, **p < 0.05, ***p < 0.01$

Models are estimated using an OLS regression. Huber-White Standard Errors In Parenthesis. Number of Observations: 351. These 351 observations correspond to the 351 local government areas in the sample. The 351 local government areas are formed using data on 1,267 firms for which ethnicity, religion and language data were available. The models in this table correspond to the Models in Table A.78. Models 7 & 9: 132 Observations dropped due to division by zero. Models 10-12: 173 Oservations dropped due to log(0) being undefined.

Table A.85: Local Government Level Estimations Of Corruption On Polarisation - Full Sample

Dependent Variable:	Corruption Index			Corruption Index (2)			Relative Corruption Index			Log Of Corruption Rents		
	1	2	3	4	5	6	7	8	9	10	11	12
Ethnic Polarisation	0.431**			0.384			-0.107			5.787***		
	(0.180)			(0.328)			(0.264)			(1.567)		
Religious Polarisation		0.267			-0.026			-0.173			4.646***	
		(0.166)			(0.287)			(0.233)			(1.583)	
Linguistic Polarisation			-1.336***			-0.596			0.196			-15.262**
			(0.482)			(1.024)			(0.744)			(6.873)
Manufacture (Share)	-0.026	-0.035	-0.015	0.046	0.049	0.093	-0.069	-0.057	-0.066	0.939	0.394	0.759
	(0.063)	(0.064)	(0.067)	(0.131)	(0.135)	(0.166)	(0.082)	(0.082)	(0.097)	(0.638)	(0.667)	(0.759)
Trade (share)	0.296**	0.298**	0.373**	0.836**	0.838**	0.744**	-0.040	-0.042	-0.066	1.809	1.863	3.085
	(0.148)	(0.148)	(0.176)	(0.390)	(0.391)	(0.367)	(0.123)	(0.123)	(0.160)	(1.207)	(1.280)	(1.893)
Constant	0.159***	0.172***	0.495***	0.154***	0.179***	0.339	0.286***	0.288***	0.248	2.590***	2.807***	6.430***
	(0.022)	(0.021)	(0.108)	(0.040)	(0.038)	(0.225)	(0.043)	(0.038)	(0.161)	(0.206)	(0.200)	(1.517)
R-squared	0.042	0.033	0.051	0.055	0.052	0.032	0.004	0.006	0.003	0.119	0.096	0.122
Observations	350	350	292	350	350	292	200	200	180	152	152	140
F-Stat	3.54	2.38	4.24	2.52	1.74	1.99	0.34	0.41	0.23	7.37	4.71	3.33
P-value	0.015	0.070	0.006	0.058	0.159	0.116	0.797	0.749	0.874	0.000	0.004	0.021

$*p < 0.10, **p < 0.05, ***p < 0.01$

. Models are estimated using an OLS regression. Huber-White Standard Errors In Parenthesis. Number of Observations: 351. These 351 observations correspond to the 351 local government areas in the sample. The 351 LGAs are generated using data on all 2,110 firms in the sample. The models in this table correspond to the Models in Table A.79, excluding the models with linguistic polarisation. Models 7 & 9: 150 Observations dropped due to division by zero. Models 10-12: 198 Oservations dropped due to log(0) being undefined.

Table A.86: Local Government Level Estimations Of Corruption On Polarisation - Reduced Sample

Dependent Variable:	Corruption Index			Corruption Index (2)			Relative Corruption Index			Log Of Corruption Rents		
	1	2	3	4	5	6	7	8	9	10	11	12
Ethnic Polarisation	0.359*			0.093			-0.090			5.692***		
	(0.200)			(0.445)			(0.321)			(2.076)		
Religious Polarisation		0.277			-0.392			-0.277			5.619***	
		(0.191)			(0.401)			(0.279)			(1.861)	
Linguistic Polarisation			-2.038***			-0.374			0.465			-11.508
			(0.623)			(1.255)			(0.843)			(7.239)
Manufacture (Share)	-0.063	-0.073	-0.057	0.008	0.033	0.010	-0.083	-0.056	-0.085	2.597**	2.158*	2.515*
	(0.064)	(0.064)	(0.062)	(0.165)	(0.168)	(0.163)	(0.110)	(0.112)	(0.111)	(1.155)	(1.166)	(1.279)
Trade (share)	0.496***	0.494***	0.513***	2.184**	2.190**	2.187**	0.117	0.114	0.110	0.807	0.826	0.946
	(0.161)	(0.161)	(0.161)	(0.975)	(0.978)	(0.972)	(0.159)	(0.159)	(0.161)	(1.192)	(1.181)	(1.224)
Constant	0.168***	0.178***	0.659***	0.183***	0.212***	0.275	0.309***	0.321***	0.198	2.442***	2.564***	5.533***
	(0.027)	(0.025)	(0.140)	(0.054)	(0.048)	(0.286)	(0.053)	(0.045)	(0.181)	(0.295)	(0.266)	(1.607)
R-squared	0.076	0.072	0.097	0.175	0.176	0.175	0.008	0.013	0.010	0.118	0.128	0.092
Observations	292	292	292	292	292	292	160	160	160	119	119	119
F-Stat	4.85	4.39	7.49	2.16	1.85	2.33	0.40	0.66	0.48	5.67	7.15	2.82
P-value	0.003	0.005	0.000	0.093	0.139	0.074	0.752	0.579	0.696	0.001	0.000	0.042

$*p < 0.10, **p < 0.05, ***p < 0.01$

. Models are estimated using an OLS regression. Huber-White Standard Errors In Parenthesis. Number of Observations: 351. These 351 observations correspond to the 351 local government areas in Nigeria. The 351 LGAs are generated using data on the 1,267 firms for which data on ethnicity, religion and language are available. The models in this table correspond to the Models in Table A.80, excluding the models with linguistic polarisation. Models 7 & 9: 132 Observations dropped due to division by zero. Models 10-12: 173 Oservations dropped due to log(0) being undefined.

Table A.87: Local Government Level Estimations Of Corruption On Linguistic Fractionalisation - Full Sample

Dependent Variable:	Corruption Index				Corruption Index (2)				Relative Corruption Index				Log Of Corruption Rents			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Avg. No. Of Languages Spoken	0.030				0.016				-0.033				0.157			
	(0.047)				(0.085)				(0.076)				(0.460)			
Linguistic Fractionalisation (Unweighted) (LF)		0.236				0.135				-0.111				4.800***		
		(0.185)				(0.324)				(0.246)				(1.693)		
Linguistic Fractionalisation $(\delta = .5)(LF_{05})$			-0.104				-0.444				-0.432				6.731***	
			(0.198)				(0.335)				(0.296)				(2.575)	
Linguistic Fractionalisation $(\delta = .05)(LF_{005})$				-0.384**				-0.819**				-0.518*				5.496**
				(0.181)				(0.341)				(0.294)				(2.158)
Manufacture (Share)	-0.011	-0.012	-0.013	-0.012	0.095	0.094	0.093	0.095	-0.068	-0.067	-0.069	-0.066	0.964	0.840	0.857	0.900
	(0.068)	(0.067)	(0.068)	(0.068)	(0.166)	(0.166)	(0.166)	(0.165)	(0.097)	(0.097)	(0.097)	(0.095)	(0.840)	(0.795)	(0.757)	(0.778)
Trade (share)	0.369**	0.367**	0.364**	0.366**	0.743**	0.742**	0.740**	0.745**	-0.070	-0.069	-0.074	-0.067	2.771	3.002	2.931	2.821
	(0.177)	(0.176)	(0.177)	(0.176)	(0.368)	(0.367)	(0.369)	(0.368)	(0.159)	(0.159)	(0.158)	(0.157)	(1.962)	(1.830)	(1.858)	(1.914)
Constant	0.122	0.052	0.247**	0.392***	0.166	0.124	0.447**	0.635***	0.367**	0.359**	0.534***	0.568***	2.723**	0.156	-0.692	0.158
	(0.109)	(0.113)	(0.115)	(0.103)	(0.203)	(0.202)	(0.202)	(0.199)	(0.179)	(0.158)	(0.177)	(0.168)	(1.087)	(1.068)	(1.480)	(1.173)
R-squared	0.036	0.040	0.035	0.043	0.032	0.032	0.034	0.040	0.004	0.004	0.013	0.017	0.056	0.124	0.139	0.110
Observations	292	292	292	292	292	292	292	292	180	180	180	180	140	140	140	140
F-Stat	1.56	2.06	1.48	2.94	1.71	1.87	1.98	3.17	0.28	0.29	0.88	1.18	1.62	4.68	4.39	4.09
P-value	0.199	0.105	0.219	0.034	0.164	0.135	0.117	0.025	0.836	0.836	0.452	0.320	0.187	0.004	0.006	0.008

$*p < 0.10, **p < 0.05, ***p < 0.01$

.Models are estimated using an OLS regression. Huber-White Standard Errors In Parenthesis.

Table A.88: Local Government Level Estimations Of Corruption On Linguistic Fractionalisation - Reduced Sample

Dependent Variable:	Corruption Index				Corruption Index (2)				Relative Corruption Index				Log Of Corruption Rents			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Avg. No. Of Languages Spoken	0.013				0.022				0.000				0.114			
	(0.054)				(0.090)				(0.088)				(0.547)			
Linguistic Fractionalisation (LF)		0.162				0.148				0.028				4.468**		
		(0.258)				(0.381)				(0.288)				(1.879)		
Linguistic Fractionalisation $(\delta = .5)(LF_{05})$			-0.273				-0.489				-0.325				7.090**	
			(0.304)				(0.444)				(0.394)				(2.783)	
Linguistic Fractionalisation $(\delta = .05)(LF_{005})$				-0.596**				-0.911**				-0.488				6.348***
				(0.263)				(0.451)				(0.407)				(2.360)
Manufacture (Share)	-0.056	-0.058	-0.054	-0.051	0.010	0.008	0.014	0.017	-0.083	-0.083	-0.081	-0.079	2.674**	2.146*	2.075*	2.263*
	(0.063)	(0.062)	(0.064)	(0.064)	(0.163)	(0.163)	(0.164)	(0.163)	(0.109)	(0.109)	(0.110)	(0.109)	(1.343)	(1.283)	(1.169)	(1.202)
Trade (share)	0.501***	0.502***	0.496***	0.500***	2.189**	2.188**	2.182**	2.188**	0.120	0.122	0.112	0.117	0.701	0.978	0.919	0.785
	(0.162)	(0.162)	(0.161)	(0.160)	(0.974)	(0.972)	(0.973)	(0.973)	(0.161)	(0.160)	(0.159)	(0.158)	(1.251)	(1.209)	(1.213)	(1.227)
Constant	0.164	0.100	0.344**	0.508***	0.140	0.103	0.457*	0.668***	0.300	0.283	0.482**	0.560**	2.774**	0.338	-0.897	-0.302
	(0.126)	(0.156)	(0.173)	(0.146)	(0.213)	(0.228)	(0.252)	(0.250)	(0.209)	(0.184)	(0.232)	(0.229)	(1.302)	(1.167)	(1.603)	(1.290)
R-squared	0.066	0.068	0.069	0.083	0.175	0.175	0.176	0.180	0.008	0.008	0.013	0.019	0.057	0.108	0.137	0.120
Observations	292	292	292	292	292	292	292	292	160	160	160	160	119	119	119	119
F-Stat	3.40	3.57	3.63	5.18	1.90	2.43	1.84	2.46	0.38	0.39	0.58	0.81	1.50	3.94	4.29	4.69
P-value	0.018	0.015	0.013	0.002	0.129	0.066	0.139	0.063	0.765	0.762	0.631	0.490	0.219	0.010	0.007	0.004

$*p < 0.10, **p < 0.05, ***p < 0.01$

.Models are estimated using an OLS regression. Huber-White Standard Errors In Parenthesis.

Table A.89: Firm Level Estimations Of Bribery On Fractionalisation

Dependent Variable:	Dummy=1 if firm admitted to bribing; 0 otherwise					
	1	2	3	4	5	6
Ethnic Fractionalisation (State Level)	0.131					
	(0.250)					
Ethnic Fractionalisation (LG Level)		0.421*				
		(0.219)				
Religious Fractionalisation (State Level)			0.764			
			(0.763)			
Religious Fractionalisation (LG Level)				0.630**		
				(0.313)		
Linguistic Fractionalisation (Unweighted) (State Level)					1.933*	
					(1.055)	
Linguistic Fractionalisation (Unweighted) (LG Level)						1.248**
						(0.631)
Employees=50-100	0.168*	0.149	0.183*	0.166	0.176*	0.160
	(0.095)	(0.136)	(0.096)	(0.136)	(0.095)	(0.137)
Employees=100-250	-0.138	-0.147	-0.162	-0.129	-0.148	-0.147
	(0.120)	(0.152)	(0.118)	(0.158)	(0.119)	(0.153)
Employees=Over 250	-0.312*	-0.313*	-0.317	-0.317**	-0.313*	-0.312*
	(0.189)	(0.162)	(0.196)	(0.162)	(0.189)	(0.160)
Trade	1.187***	1.167***	1.169***	1.164***	1.180***	1.175***
	(0.232)	(0.175)	(0.256)	(0.186)	(0.231)	(0.177)
Foreign ownership $\geq$ 25%	-0.063	-0.075	-0.076	-0.098	-0.086	-0.078
	(0.153)	(0.143)	(0.171)	(0.139)	(0.152)	(0.143)
Dummy=1 if company using security service	0.526	0.512	0.606	0.639	0.458	0.464
	(0.461)	(0.608)	(0.450)	(0.636)	(0.452)	(0.624)
Log of age	-0.051	-0.049	-0.050	-0.046	-0.049	-0.044
	(0.064)	(0.045)	(0.060)	(0.047)	(0.064)	(0.045)
Constant	-1.085	-1.169*	-1.341***	-1.296**	-2.231**	-1.772**
	(0.728)	(0.608)	(0.493)	(0.639)	(0.881)	(0.770)
Pseudo R-squared	0.050	0.056	0.058	0.057	0.057	0.054
Observations	1255	1255	1255	1255	1255	1255
Chi-Squared	84.88	76.97	101.96	99.72	84.68	74.90
P-value	0.000	0.000	0.000	0.000	0.000	0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$

. Coefficients are from a Probit Estimation. Region (Models 1, 3 & 5) & Local Government Area (Models 2, 4 & 6) Clustered Standard Errors In Parenthesis. 12 Observations dropped due to log(0) being undefined. The base category for "Employees" is "Less than 50".

Table A.90: Firm Level Estimations Of Bribery On Polarisation

Dependent Variable:	Dummy=1 if firm admitted to bribing; 0 otherwise					
	1	2	3	4	5	6
Ethnic Polarisation (State Level)	3.035					
	(2.101)					
Ethnic Polarisation (LG Level)		2.175***				
		(0.768)				
Religious Polarisation (State Level)			1.552			
			(1.559)			
Religious Polarisation (LG Level)				1.300**		
				(0.649)		
Linguistic Polarisation (State Level)					-4.377	
					(2.747)	
Linguistic Polarisation (LG Level)						-5.178***
						(1.879)
Employees=50-100	0.152	0.120	0.183*	0.166	0.175*	0.163
	(0.101)	(0.133)	(0.096)	(0.136)	(0.094)	(0.138)
Employees=100-250	-0.171	-0.158	-0.162	-0.129	-0.144	-0.149
	(0.114)	(0.153)	(0.118)	(0.158)	(0.120)	(0.154)
Employees=Over 250	-0.305	-0.312*	-0.317	-0.320**	-0.316*	-0.313*
	(0.196)	(0.164)	(0.196)	(0.162)	(0.188)	(0.160)
Trade	1.166***	1.138***	1.169***	1.164***	1.186***	1.175***
	(0.252)	(0.178)	(0.256)	(0.186)	(0.231)	(0.178)
Foreign ownership $\geq$ 25%	-0.108	-0.082	-0.076	-0.096	-0.077	-0.086
	(0.170)	(0.143)	(0.171)	(0.139)	(0.150)	(0.143)
Dummy=1 if company using security service	0.429	0.482	0.609	0.640	0.483	0.451
	(0.500)	(0.597)	(0.448)	(0.637)	(0.447)	(0.627)
Log of age	-0.039	-0.047	-0.050	-0.046	-0.050	-0.043
	(0.058)	(0.045)	(0.061)	(0.047)	(0.064)	(0.045)
Constant	-1.316**	-1.246**	-1.341***	-1.300**	-0.091	0.133
	(0.568)	(0.596)	(0.491)	(0.640)	(0.873)	(0.718)
Pseudo R-squared	0.062	0.063	0.058	0.057	0.054	0.058
Observations	1255	1255	1255	1255	1255	1255
Chi-Squared	92.41	85.54	100.57	100.39	84.08	76.45
P-value	0.000	0.000	0.000	0.000	0.000	0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$

Coefficients are from a Probit Estimation. Region (Models 1, 3 & 5) & Local Government Area (Models 2, 4 & 6) Clustered Standard Errors In Parenthesis. 12 Observations dropped due to log(0) being undefined.

Table A.91: Firm Level Estimations Of Bribery On Linguistic Fractionalisation

Dependent Variable:	Dummy=1 if firm admitted to bribing; 0 otherwise					
	1	2	3	4	5	6
Linguistic Fractionalisation (State Level) (Unweighted)	1.933*					
	(1.055)					
Linguistic Fractionalisation (LG Level) (Unweighted)		1.248**				
		(0.631)				
Linguistic Fractionalisation (State Level) ($\delta = .5$)(LF_{05})			-0.116			
			(1.636)			
Linguistic Fractionalisation (LG Level) ($\delta = .5$)(LF_{05})				-0.150		
				(0.710)		
Linguistic Fractionalisation (State Level) ($\delta = .05$)(LF_{005})					-1.485	
					(1.673)	
Linguistic Fractionalisation (LG Level) ($\delta = .05$)(LF_{005})						-1.444**
						(0.731)
Employees=50-100	0.176*	0.160	0.165*	0.166	0.162*	0.167
	(0.095)	(0.137)	(0.094)	(0.138)	(0.095)	(0.136)
Employees=100-250	-0.148	-0.147	-0.132	-0.131	-0.124	-0.124
	(0.119)	(0.153)	(0.124)	(0.154)	(0.125)	(0.154)
Employees=Over 250	-0.313*	-0.312*	-0.313*	-0.313**	-0.308	-0.309*
	(0.189)	(0.160)	(0.190)	(0.159)	(0.193)	(0.159)
Trade	1.180***	1.175***	1.185***	1.186***	1.185***	1.189***
	(0.231)	(0.177)	(0.229)	(0.176)	(0.240)	(0.178)
Foreign ownership $\geq$ 25%	-0.086	-0.078	-0.055	-0.055	-0.051	-0.049
	(0.152)	(0.143)	(0.151)	(0.142)	(0.160)	(0.144)
Dummy=1 if company using security service	0.458	0.464	0.549	0.551	0.601	0.594
	(0.452)	(0.624)	(0.428)	(0.618)	(0.433)	(0.621)
Log of age	-0.049	-0.044	-0.052	-0.052	-0.049	-0.055
	(0.064)	(0.045)	(0.063)	(0.046)	(0.061)	(0.047)
Constant	-2.231**	-1.772**	-0.988	-0.970	-0.295	-0.312
	(0.881)	(0.770)	(1.486)	(0.775)	(1.505)	(0.760)
Pseudo R-squared	0.057	0.054	0.050	0.050	0.053	0.053
Observations	1255	1255	1255	1255	1255	1255
Chi-Squared	84.68	74.90	103.53	74.19	149.37	75.37
P-value	0.000	0.000	0.000	0.000	0.000	0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$

.Coefficients are from a Probit Estimation. Region (Models 1, 3 & 5) & Local Government Area (Models 2, 4 & 6) Clustered Standard Errors In Parenthesis. 12 Observations dropped due to log(0) being undefined.

Table A.92: South-West Zone: Firm Level Probit Estimations Of Bribery On Matching, With Regional Dummies

Dependent Variable:	Dummy=1 if firm admitted to bribing; 0 otherwise					
	1-SW	2-SW	3-SW	4-SW	5-SW	6-SW
Similar Ethnicity	-0.508***	-0.452***	-0.424***	-0.003	0.013	0.036
	(0.131)	(0.134)	(0.137)	(0.146)	(0.148)	(0.151)
Trade		0.768***	0.839***		0.427*	0.501**
		(0.205)	(0.211)		(0.228)	(0.234)
Employees=50-100			0.030			-0.111
			(0.191)			(0.203)
Employees=100-250			-0.256			-0.415*
			(0.213)			(0.236)
Employees=Over 250			-0.402*			-0.416*
			(0.223)			(0.248)
Reg==Ekiti				-1.976***	-1.993***	-2.065***
				(0.462)	(0.438)	(0.434)
Reg==Ogun				-2.096***	-2.035***	-2.004***
				(0.444)	(0.445)	(0.445)
Reg==Ondo				-1.994***	-1.929***	-1.935***
				(0.463)	(0.465)	(0.466)
Reg==Osun				-1.840***	-1.819***	-1.877***
				(0.347)	(0.342)	(0.342)
Reg==Oyo				-1.142***	-1.078***	-1.124***
				(0.246)	(0.248)	(0.247)
Constant	0.065	-0.057	-0.027	0.210*	0.133	0.206
	(0.108)	(0.114)	(0.123)	(0.112)	(0.120)	(0.131)
Pseudo R-squared	0.026	0.051	0.058	0.233	0.241	0.249
Observations	439	439	439	439	439	439
Chi-Squared	14.96	27.85	32.10	94.71	110.04	119.31
P-value	0.000	0.000	0.000	0.000	0.000	0.000

$*p < 0.10, **p < 0.05, ***p < 0.01$

. Heteroskedastic Robust Standard Errors In Parenthesis. Similar Ethnicity: Dummy=1 if ethnicity of manager matches ethnicity of LGA Chair, 0 otherwise. Unit of observation is the firm. Similar Ethnicity is calculated using manager data collected by the Author. All other variables are calculated using data from the NBS/EFCC survey.

References

Abeler, J., Becker, A. & Falk, A., (2012). Truth-Telling: A Representative Assessment. Institute For The Study Of Labour. Discussion Paper No. 6919.

Abrevaya, J. & Hausman, J.A., (1999). Semiparametric Estimation With Mismeasured Dependent Variables: Application To Duration Models For Unemployment Spells. *Annals Of Economics And Statistics*, No. 55/56. Panel Data Econometrics (sep. - Dec., 1999), pp. 243-275.

Ades, A. and Di Tella, R. (1999) Rents, Competition, And Corruption. *American Economic Review*. 89, 4, pp. 982-993.

Aghion, P., Alesina, A. & Trebbi, F., (2004). Polarization And Institutional Choice. Unpublished.

Aidt, T. S., (2003). Economic Analysis Of Corruption: A Survey. *The Economic Journal*. 113.

Aitken, S.S. & Bonneville, L., (1980). A General Taxpayer Opinion Survey, Prepared For Office Of Planning And Research, Internal Revenue Service, Washington: CSR, Inc., 1980.

Alesina, A.; Devleeschauwer, A.; Easterly, W.; Kurlat, S. & Wacziarg, R., (2003). Fractionalization. *Journal Of Economic Growth*. 8, pp. 155-194.

Alesina, A. & La Ferrara, E., (2000). Participation In Heterogeneous Communities. *Quarterly Journal Of Economics*. 115, 3, pp. 847-904.

Alesina, A. & La Ferrara, E., (2005). Ethnic Diversity And Economic Performance. *Journal Of Economic Literature*. Vol. 43, pp. 762800.

Alesina, A., Spolaore, E. & Wacziarg, R., (2000).Economic Integration And Political Disintegration. *American Economic Review*. 90

Almeida, R. & Carneiro, P., (2005). Enforcement Of Labor Regulation, Informal Labor, And Firm Performance. World Bank Policy Research Working Paper 3756.

Aluko, J.O. (2006) Corruption In The Local Government System In Nigeria. Editions Africa. Nigeria.

ANGELUCCI, A., DE GIORGI, G., RANGEL, M.A. & RASUL, I., (2010). Family Networks And School Enrolment: Evidence From A Randomized Social Experiment. *Journal of Public Economics*. Vol. 94, pp. 197-221.

ANGELUCCI, A., DE GIORGI, G. & RASUL, I., (2012). Resource Pooling Within Family Networks: Insurance And Investment. Working Paper.

ARTÍS, M., AYUSO, M. & GUILLÉN, M., (2002). Detection Of Automobile Insurance Fraud With Discrete Choice Models And Misclassified Claims. *The Journal Of Risk And Insurance*. Vol. 69, No. 3, pp. 325-340.

AZFAR, O. & MURRELL, P., (2009). Identifying Reticent Respondents: Assessing The Quality Of Survey Data On Corruption And Values. *Economic Development and Cultural Change*. 57, 2, pp. 387-411.

BANDIERA, O., BARANKAY, I. & RASUL, I., (2005). Cooperation In Collective Action. *Economics Of Transition*. Vol. 13 (3), pp. 473-498.

BANDIERA, O., BARANKAY, I. & RASUL, I., (2009). Social Connections And Incentives In The Workplace: Evidence From Personnel Data. *Econometrica*, 77 (4), pp. 1047-1094.

BANDIERA, O., BARANKAY, I. & RASUL, I., (2010). Social Incentives In The Workplace. *Review Of Economic Studies*. 77 (2), pp. 417-458.

BANDIERA, O., PRAT, A. & SADUN, (2013). Managing Firms In An Emerging Economy: Evidence From The Time Use Of Indian CEOs. Available Online: http://www.bus.umich.edu/academics/departments/be/pdf/Bandiera%20CEOIndiaTimeUse.pdf Accessed 2nd September 2013.

BARDHAN, P. (1997). Corruption And Development: A Review Of Issues. *Journal Of Economic Literature*. 1320-1346

BARR, A. & SERRA, D., (2010). Corruption And Culture: An Experimental Analysis. *Journal Of Public Economics*. 94, 11-12. pp. 862-869.

BARRETT, D.B., KURIAN, G.T., & JOHNSON, T.M., (2001). World Christian Encyclopedia: A Comparative Survey Of Churches And Religions. Second Edition. Oxford: Oxford University Press.

BAULCH, B., PHAM, H.T. & REILLY, B., (2012). Decomposing The Ethnic Gap In Rural Vietnam: 1993 to 2004. Oxford Development Studies, 40 (1). pp. 87-117.

Bécares, L., Nazroo, J. & Stafford, M., (2009). The Buffering Effects Of Ethnic Density On Experienced Racism And Health. *Health & Place*.

Bécares, L., Stafford, M., Laurence, J. & Nazroo, J., (2011). Composition, Concentration And Deprivation: Exploring Their Association With Social Cohesion Among Different Ethnic Groups In The UK. *Urban Studies*. Available Online: http://usj.sagepub.com/content/early/2011/01/15/0042098010391295. Accessed on 7th August 2013.

Becker, G.S., (1968). Crime And Punishment: An Economic Approach. *Journal Of Political Economy*, Vol. 76, No. 2, pp. 169-217.

Becker, S.O., Egger, P.H. and Seidel, T. (2009). Common Political Culture: Evidence On Regional Corruption Contagion. *European Journal Of Political Economy*. 25, pp. 300-310.

Belot, M. & van de Ven, J., (2009). Friendships And Favouritism On The Schoolground - A Framed Field Experiment. Unpublished Manuscript. Oxford University.

Berg, J., Dickhaut, J. & McCabe, K., (1995). Trust, Reciprocity, And Social History. *Games And Economic Behaviour*. 10, pp. 122-142.

Bertrand, M. & Mullainathan, S., (2001). Do People Mean What They Say? Implications For Subjective Survey Data. *AEA Papers And Proceedings*, Vol. 91, No. 2, pp. 67-72.

Bliss, C. & Di Tella, R. (1997). Does Competition Kill Corruption. *Journal Of Political Economy*. Vol. 105 (5): 1001-1023.

Bluedorn, J., (2001). Can Democracy Help? Growth And Ethnic Divisions. *Economics Letters*. 70, pp.121-126.

Bossert, W., D'Ambrosio, C. & La Ferrara, E., (2011). A Generalized Index Of Fractionalization. *Economica*. Vol. 78, pp. 723-750.

Bound, J., Brown, C. & Mathiowetz, N., (2001). Measurement Error In Survey Data. In: Heckman, J.J. & Leamer, E., (Eds.) Handbook Of Econometrics, Volume 5. Elsevier Science B.V. The Netherlands.

Bradburn, N.M.,Sudman, S.,Blair, E., Locander, W., Miles, C., and Singer, E., (1991). Improving Interview Method And Questionnaire Design: Response Effects To Threatening Questions In Survey Research. University Microfilms International.

Brandts, J. & Solà, C., (2010). Personal Relations And Their Effect On Behaviour In An Organisational Setting: An Experimental Study. *Journal Of Economic Behaviour And Organisation.* 73 (2), pp.246-253.

Brinkhoff, T., (2013). City Population. Available Online: http://www.citypopulation.de. Date Accessed: 14th February 2013.

Brück, T., Justino, P., Verwimp, P. & Avdeenko, A., (2010). Identifying Conflict And Violence In Micro-Level Surveys. *IZA Discussion Paper No. 5067*

Burgess, R., Jedwab, R., Miguel, E., Morjaria, A. & i Miquel, G.P., (2011). Ethnic Favouritism. Working Paper

Cameron, A.C. and Trivedi, P.K. (2005). Microeconometrics: Methods And Applications. Cambridge University Press. Cambridge.

Cameron, A.C. and Trivedi, P.K. (2009). Microeconometrics Using Stata. Stata Press. USA.

Carlin, W. & Seabright, P., (2007). Bring Me Sunshine: Which Parts Of The Business Environment Should Public Policy Try To Fix? Annual World Bank Conference On Development Economics. Bled, Slovenia.

Caselli, F. & Coleman, J., (2002). On The Theory Of Ethnic Conflict. Unpublished. Harvard University.

Central Intelligence Agency, (2013). The World Factbook. "https://www.cia.gov/library/publications/the-world-factbook/geos/ni.html". Date Accessed: 16th January 2013.

Cerqueti, R., Coppier, R. & Piga, G., (2012). Corruption, Growth And Ethnic Fractionalization: A Theoretical Model. *Journal Of Economics.* Vol. 106, pp. 153-181

Clarke, G.R.G., (2011). How Petty Is Petty Corruption? Evidence From Firm Surveys In Africa. *World Development.* 39, 7, pp. 1122-1132.

Clarke, G., (2012). Lying About Firm performance: Evidence From A Survey In Nigeria. MPRA Paper No. 3538.

Clausen, B., Kraay, A. & Murrell, P., (2010). Does Respondent Reticence Affect The Results Of Corruption Surveys? Evidence From The World Bank Enterprise Survey For Nigeria.World Bank.

Clots-Figueras, I. & Masella, P., (2013). Education, Language and Identity. *Economic Journal*, 123 (570).

Coldman, A., Braun, T. & Gallagher, R. P., (1988). The Classification Of Ethnic Status Using Name Information. *Journal Of Epidemiology And Community Health*, Vol. 42, pp. 390-395.

Collier, P., (2001). Implications Of Ethnic Diversity. *Economic Policy*. Volume 16, Issue 32. pp. 127-166.

Collier, P. & Hoeffler, A., (2002). Greed And Grievances. World Bank. DECRG Working Paper.

Cruyff, M.J.L.F., Van Den Hout, A., & Van Der Heijden, P.G.M., (2008). The Analysis Of Randomized-Response Sum Score Variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 70, pp. 21-30.

Delavallade, C. (2012), What Drives Corruption? Evidence from North African Firms, *Journal Of African Economies*. Vol. 21 (4): pp.499-547

De Rosa, D., Gooroochurn, N. & Görg, H., (2010). Corruption And Productivity: Firm-Level Evidence From The BEEPS Survey. Kiel Working Papers. Kiel Institute For The World Economy.

Desmet, K., Ortuño-Ortín, I. & Wacziarg, R., (2009). The Political Economy Of Ethnolinguistic Cleavages. *NBER Working Paper*. No. 15360. Available Online: http://www.nber.org/papers/w15360.Accessed: 17th February 2013.

Desmet, K., Ortuño-Ortín, I. & Weber, S., (2009). Linguistic Diversity And Redistribution. *Journal Of The European Economic Association*. Volume 7, pp. 1291-1318.

Dickens R. & McKnight, A., (2008). Assimilation Of Migrants Into The British Labour Market. *Working Paper*. Centre for Economic Performance, LSE, London.

Dollar, D., Hallward-Driemeier, M. & Mengistae, T., (2005). Investment Climate And Firm Performance In Developing Economies. *Economic Development And Cultural Change*. 54, 1, pp. 1-31.

Do, Nguyen & Tran, (2013). One Mandarin Benefits The Whole Clan: Hometown Favouritism In An Authoritarian Regime. Working Paper.

Easterly, W., (2001). Can Institutions Resolve Ethnic Conflict? *Economic Development And Cultural Change*. 49 (4) pp. 687-706.

EASTERLY, W. & LEVINE, R., (1997). Africa's Growth Tragedy: Policies And Ethnic Divisions. *Quarterly Journal Of Economics*. 112, pp. 1203-1250.

EHRENBERG, R.G., GOLDHABER, D.D. & BREWER, D.J., (1995). Do Teachers' race, gender, and ethnicity matter? Evidence from the National Educational Longitudinal Study Of 1988. *Industrial And Labour Relations Review*, 48, 3, pp. 547-561.

ENRICK, N.L., (1963). A Pilot Study Of Income Tax Consciousness. National Tax Journal. 16, (2), pp. 169-173.

ERARD, B. and FEINSTEIN, J.(1994), Honesty and evasion in the tax compliance game. *Rand Journal Of Economics*. 25,1, pp. 1-19

ERCOLANI, M.G., (2006). UK Employees' Sickness Absence, 1984-2005. University Of Birmingham.

ESTEBAN, J., MAYORAL, L. & RAY, D., (2012). Ethnicity And Conflict: An Empirical Study. *American Economic Review*. Vol. 102, Issue 4, pp. 1310-1342.

FAFCHAMPS, M., (2000). Ethnicity And Credit In African Manufacturing. *Journal Of Development Economics*, Vol. 61 (1), pp. 205-235.

FAFCHAMPS, M., (2003). Ethnicity And Networks In African Trade. *Contributions In Economic Analysis & Policy*, 2 (1).

FALARIS, E.M., (2007). Misclassification Of The Dependent Variable In Binary Choice Models: Evidence From Five Latin American Countries. Working Paper No. 2007-05, University of Delaware, Department of Economics.

FAN, C.S., LIN, C. and TREISMAN, D., (2010). Embezzlement Versus Bribery. *National Bureau of Economic Research Working Paper 16542*.

FARKAS, G., GROBE, R.P., SHEEHAN, D. & SHUAN, Y., (1990). Cultural Resources And School Success: Gender, Ethnicity, And Poverty Groups Within An Urban District. *American Sociological Review*, 55 (1), pp. 127-142.

FEARON, J., (2003). Ethnic And Cultural Diversity By Country. *Journal Of Economic Growth*. Vol. 8, No. 2, pp. 195-222.

FERSHTMAN, C., GNEEZY, U. & VERBOVEN, F., (2005). Discrimination And Nepotism: The Efficiency Of The Anonymity Rule. *Journal Of Legal Studies*. 34 (2): pp. 371-394.

Fetchenhauer, D. & Dunning, D., 2009. Do People Trust Too Much Or Too Little? *Journal Of Economic Psychology.* Vol. 30, Issue 3, pp. 263-276.

Fischbacher, U. & Heusi, F., (2008). Lies In Disguise: An Experimental Study On Cheating. TWI Research Paper, 40.

Fisman, R. & Miguel, E., (2007). Corruption, Norms, And Legal Enforcement: Evidence From Diplomatic Parking Tickets. *Journal of Political Economy.* 115, 6, pp. 1020-1048.

Foreign Corrupt Practices Act (1977). Unlawful Corporate Payment Act Of 1977. *House Of Representatives.* Report No. 95-640.

Fowler, F.J., (1995). Improving Survey Questions: Design And Evaluation. *Applied Social Research Methids Series.* 38.

Fox, J.A.; & Tracy, P.E., (1986). Randomized Response. A Method For Sensitive Surveys. Sage Publications, Inc. USA.

Frank, B. & Schulze, G. G., (2000). Does Economics Make Citizens Corrupt? *Journal Of Economic Behaviour & Organization.* Vol. 43, p. 101-113.

Friedland, N., Maital, S. & Rutenberg A., (1978). A Simulation Study Of Income Tax Evasion. *Journal Of Public Economics.* 10, pp. 107-116.

Fukuyama, F., 1995. Trust: The Social Virtues And The Creation Of Prosperity. Free Press, New York.

Garcia-Montalvo, J. & Reynal-Querol, M., (2002). Why Ethnic Fractionalization? Polarization, Ethnic Conflict And Growth. Unpublished. Universitat Pompeu Fabra.

Golden, M.A. & Picci, L., (2005). Proposal For A New Measure Of Corruption, Illustrated With Italian Data. *Economics & Politics.* 17, 1, pp. 37-75.

Good, P.I. & Hardin, J.W., (2006). Common Errors In Statistics (And How To Avoid Them). John Wiley & Sons. Canada.

Greenberg, J.H., (1956). The Measurement Of Linguistic Diversity. *Language.* Vol. 32, No. 1, pp. 109-115.

Greenberg, B.G., Abul-Ela, A.L.A, Simmons, W.R. & Horvitz, D.G., (1969). The Unrelated Question Randomised Response Model: Theoretical Framework. *Journal Of The American Statistical Association.* pp.520-539.

GREENE, W.H.(2003), Econometric Analysis. Fifth Edition. Pearson Education Inc. USA.

GREENE, W.H., (2008). The Econometric Approach To Efficiency Analysis. *The Measurement Of Productive Efficiency And Productivity Growth*. Oxford University Press. pp.92-250.

GREIF, A., (1993). Contract Enforceability And Economic Institutions In Early Trade: The Maghribi Traders' Coalition. *American Economic Review*. 83, 3, pp.525-548.

GROGGER, J., (2011). Speech Patterns And Racial Wage Inequality. *Journal of Human Resources*, 46(1), pp. 1-25.

HARDIN, J.W. & HILBE, J., (2007). Generalized Linear Models And Extensions. Stata Corp.

HAUSMAN, J.A.(1978), Specification Tests In Econometrics. *Econometrica*. 46, pp. 1251-1271.

HAUSMAN, J., (2001). Mismeasured Variables In Econometric Analysis: Problems From The Right And Problems From The Left. *The Journal Of Economic Perspectives*, Vol. 15, No. 4, pp. 57-67.

HAUSMAN, J.A., ABREVAYA, J. & SCOTT-MORTON, F.M., (1998). Misclassification Of The Dependent Variable In A Discrete-Response Setting. *Journal Of Econometrics*, 87, pp. 239-269.

HAUSMAN, J.A. and MCFADDEN, D.(1984), Specification Tests For The Multinomial Logit Model, *Econometrica*. 52, pp. 1219-1240

HECKMAN, J.(1979) Sample Selection Bias As A Specification Error. *Econometrica*. 47,1, pp.153-161.

HELLMAN, J.S., JONES, G., and KAUFMANN, D., (2000) September. "Seize the state, seize the day": state capture, corruption, and influence in transition. Policy research working paper series 2444. World Bank.

HOEFLICH, M.H., (1983). Of Reason, Gamesmanship, And Taxes: A Jurisprudential And Games Theoretical Approach To The Problem Of Voluntary Compliance. *The American Journal Of Tax Policy*. 2, 9, pp. 9-88.

HÖFLER, M., (2005). The Effect Of Misclassification On The Estimation Of Association: A Review. *International Journal Of Methods In Psychiatric Research*, Vol. 14, No. 2, pp. 92-101.

HOLM, H. & NYSTEDT, P., 2008. Trust In Surveys And Games - A Methodological Contribution On The Influence Of Money And Location. *Journal Of Economic Psychology.* Vol. 29, Issue 4, pp. 522-542.

HOROWITZ, D.L., (2001). The Deadly Ethnic Riot. Berkeley, CA: University Of California Press.

HORVITZ, D.G., SHAH, B.V. and SIMMONS, W.R., (1967). The Unrelated Question Randomised Response Model.65, 72.

HUMAN RIGHTS WATCH, (2010). "Everyone's in on the Game": Corruption and Human Rights Abuses by the Nigeria Police Force. Available Online: http://www.hrw.org/en/reports/2010/08/17/everyone-s-game-0. Accessed: 20th September 2012.

HUMPHREYS, M., POSNER, D.N. & WEINSTEIN, J.M., (2002). Ethnic Identity, Collective Action, And Conflict: An Experimental Approach. Unpublished. Harvard University And UCLA.

HUNT, J., (2004). Trust And Bribery. The Role Of The Quid Pro Quo And The Link With Crime. *NBER Working Paper Series.* Working Paper 10510. NBER. Cambridge, MA.

HUNT, J., (2006). Why Are Some Public Officials More Corrupt Than Others? In Susan Rose-Ackerman (ed.) International Handbook on the Economics of Corruption. Edward Elgar. Northampton, M.A.

HUNT, J., (2007). How Corruption Hits People When They Are Down. *Journal Of Development Economics.* Volume 84, Issue 2, pp. 574-589.

HUNT, J., (2008) The Contrasting Distributional Effects Of Police And Judicial Corruption. Annual Bank Conference On Development Economics.

HUNT, J. & LASZLO, S., (2012). Is Bribery Really Regressive? Bribery's Costs, Benefits and Mechanisms. *World Development.* Vol. 40, Issue: 2, pp. 355-372.

HUNTINGDON, S.P., (1968). Political Order In Changing Societies. Yale University. USA.

IAROSSI, G., (2006). The Power Of Survey Design: A User's Guide For Managing Surveys, Interpreting Results, And Influencing Respondents. World Bank

IAROSSI, G. & CLARKE, G.R.G, (2011). An Assessment Of The Investment Climate In 26 States. The World Bank.

IAROSSI, G, MOUSLEY P. & RADWAN, I., (2009). An Assessment Of The Investment Climate In Nigeria. Directions In Development: Private Sector Development. The World Bank.

INDEPENDENT NATIONAL ELECTORAL COMMISSION, (2010). Registration Area; Local Government Area Analysis - Nationwide. INEC. Available Online: http://www.inecnigeria.org/wp-content/uploads/downloads/2010/11/RA-LGA-ANALYSIS-NATIONWIDE.pdf. Accessed: 18th February 2013.

ISAKSSON, A., (2013). Corruption Along Ethnic Lines: A Study Of Individual Corruption Experiences In 17 African Countries. University Of Gothenburg Working Paper No. 571.

JAIN, A.K., (2001). Corruption: A Review.*Journal Of Economic Surveys*. Vol. 15, No.1.

JENSEN, N.M., LI, Q. and RAHMAN, A., (2010). Understanding Corruption And Firm Responses In Cross-National Firm-Level Surveys.*Journal Of International Business Studies*, 41, 9, 1481-1504.

JENSEN, N. and RAHMAN, A., (2011). The Silence Of Corruption: Identifying Underreporting Of Business Corruption Through Randomised Response Techniques. World Bank Policy Research Working Paper No. 5696.

JOHNSON, SIMON, KAUFMANN, D., MCMILLAN, J., and WOODRUFF, C. 2000. Why do firms hide? Bribes and unofficial activity after communism. *Journal of Public Economics* 76 (3), pp. 495-520.

KANTOR, P., (2003). Women's Empowerment Through Home-Based Work: Evidence From India. *Development And Change*. Vol. 34, Issue 3, pp. 425-445.

KASFIR, N., (1976). The Shrinking Political Arena: Participation And Ethnicity In African Politics. Berkeley And Los Angeles: University Of California Press.

KAUFMANN, D., KRAAY, A., and MASTRUZZI, M., (2005). Governance Matters IV: Governance Indicators For 1996-2004. Mimeo, World Bank.

KAUFMANN, D., KRAAY, A., and MASTRUZZI, M., (2006) Measuring Corruption: Myths and Realities. World Bank.

KERKVLIET, J., (1994a). Cheating By Economics Students: A Comparison Of Survey Results. *The Journal Of Economics Education*. Vol. 25, No. 2, pp. 121-133.

KERKVLIET, J., (1994b). Estimating A Logit Model With Randomized Data: The Case Of Cocaine Use. *Australian Journal Of Statistics.* 36, 1, 9-20.

KLEIN, S., LE, V. & HAMILTON, L., (2001). Does Matching Student And Teacher Racial/Ethnic Group Improve Math Scores? Rand DOcument.

KRAAY, A. & MURRELL, P., (2013). Misunderestimating Corruption. World Bank Policy Research Working Paper No. 6488.

LA FERRARA, E., (2003). Kin Groups And Reciprocity: A Model Of Credit Transactions In Ghana. *American Economic Review.*

LAMBSDORFF, J.G., (1998) An Empirical Investigation Of Bribery In International Trade. The European Journal Of Development Research, 10, 1, pp.40-59.

LAMBSDORFF, J.G., (1999). The Transparency International Corruption Perceptions Index 1999: Framework Document. Transparency International.

LAMBSDORFF, J.G., (2003). Background Paper To The 2003 Corruption Perceptions Index. Mimeo: Transparency International, September 2003.

LAMBSDORFF, J.G., (2007). The Methodology Of The Corruption Perceptions Index 2007. Transparency International.

LAWS OF THE FEDERATION OF NIGERIA, (1990). Volume 5 Chapter 27: CRIMINAL CODE ACT, Chapter 12: Corruption and Abuse of Office.

LAWS OF THE FEDERATION OF NIGERIA. Taxes And Levies (Approved List For Collection) Decree No. 21 of 1998. Schedule: Part I; II; & III.

LAZEAR, E., (1999a). Culture And Language. *Journal Of Political Economy, Supplement.* pp. 95-125.

LAZEAR, E., (1999b). Globalization And The Market For Team-Mates. *Economic Journal.* 109, pp.15-40.

LEITE, C. and WEIDMANN, J., (2002): Does mother nature corrupt? In: G.T. Abed and S. Gupta, eds. (2002) Natural resources, corruption, and economic growth. *Governance, corruption, and economic performance.* International Monetary Fund. pp. 159-196.

LENSVELT-MULDERS, G.J.L.M., HOX, J.J., and HEIJDEN, P.G.M., (2005a). How To Improve The Efficiency Of Randomised Response Designs. *Quality And Quantity.* 39,3, pp. 253-265.

LENSVELT-MULDERS, G.J.L.M., HOX, J.J., VAN DER HEIJDEN, P.G.M. and MAAS, C.J.M., (2005b). Meta-analysis of randomized response research. *Sociological Methods & Research.* 33(3), pp. 319-348.

LIBRARY OF CONGRESS, (1947). Procedure For Military Executions. Department Of The Army.

LIN, T. & SCHMIDT, P. (1984). A Test Of The Tobit Specification Versus An Alternative Suggested By Cragg. *The Review Of Economics And Statistics.* Vol. 66, No. 1, pp.174-177.

LOCANDER, W.,SIDMAN, S. and BRADBURN, N., (1976). An Investigation Of Interview Method, Threat, And Response Distortion. *Journal Of The American Statistical Association*,pp. 269-275

LOFTUS, E.F. & ZANNI, G., (1975). Eyewitness Testimony: The Influence Of The Wording Of A Question. *Bulletin Of The Psychonomic Society.* 5, 1, pp. 86-88.

MADDALA, G.S., (1983). Limited-Dependent And Qualitative Variables In Econometrics. Cambridge University Press. Cambridge.

MAGDER, L.S. & HUGHES, J.P., (1997). Logistic Regression When The Outcome Is Measured With Uncertainty. *American Journal Of Epidemology*, Vol. 146. No. 2.

MASELLA, P, (2013). National Identity And Ethnic Diversity. *Journal of Population Economics*, 26 (2). pp. 437-454.

MAURO, P., (1995). Corruption And Growth. *The Quarterly Journal of Economics.* Vol. 110, No.3, pp. 681-712.

MCCROHAN, K., (1982). The Use Of Survey Research To Estimate Trends In Non-Compliance With Federal Income Taxes. *Journal Of Economic Psychology*, 2, pp. 231-240.

MCFADDEN, D., 1984. Econometric Analysis Of Wualitative Response Models. In: Griliches, Z., Intriligator, M.D.(Eds.), Handbook Of Econometrics, vol. 4. North-Holland. Amsterdam.

MÉNDEZ, F. and SEPÚLVEDA,F., (2010). What Do We Talk About When We Talk About Corruption? *Journal Of Law, Economics, and Organization*, 26, 3, p.493

Michalopoulos, S. & Papaioannou, E., (2013). Pre-Colonial Ethnic Institutions and Contemporary African Development. *Econometrica*. Volume 81, Issue 1, pp.113-152.

Montalvo, J. & Reynal-Querol, M., (2005a). Ethnic Polarization, Potential Conflict, And Civil Wars. *The American Economic Review*. Vol. 95, No. 3, pp. 796-816.

Montalvo, J. & Reynal-Querol, M., (2005b). Ethnic Diversity And Development. *Journal Of Development Economics*. 76, pp.293-323.

Moradi, A. & Baten, J., (2005). Inequality In Sub-Saharan Africa: New Data And New Insights From Anthropometric Estimates. World Development (8). pp. 1233-1265.

Moser, C. A. & Kalton, G., (1971). Survey Methods In Social Investigation. Heinemann Educational Book Limited. London.

Ndaguba, I., (1995). Not In Our Character. The Nation. Nigerian Television Authority. Nigeria.

Nega, F., Mathijs, E., Deckers, J. & Tollens, E., (2009). Gender, Social Capital And Empowerment In Northern Ethiopia. MPRA Paper No. 24629.

Nicoll, A., Bassett, K. & Ulijaszek, S.J., (1986). What's In A Name? Accuracy Of Using Surnames And Forenames In Ascribing Asian Ethnic Identity In English Populations. *Journal Of Epidemiology And Community Health*. Vol. 40, No. 4, pp.364-368.

Nigerian Investment Promotion Commission, (). Taxation And Fiscal Regulations In Nigeria. Mimeo.

Oguzhan, C.D., (2008). Ethnic And Religious Diversity And Corruption. *Economics Letters*. Volume 99, Issue 1, pp.98-102.

Olatunji, O.C., (2009). A Review Of Value Added Tax (VAT) Administration In Nigeria. *International Business Management*. 3, 4, pp. 61-68.

Olken, B., (2007). Monitoring Corruption: Evidence From A Field Experiment In Indonesia. *Journal Of Political Economy*. Vol. 115, pp.200-249.

Olken, B., (2009). Corruption Perceptions Vs. Corruption Reality. *Journal Of Public Economics*, Vol. 93, Issues 7-8, pp. 950-964.

OLKEN, B. & BARRON, P., (2009). The Simple Economics Of Extortion: Evidence From Trucking In Aceh. *Journal Of Political Economy.* University Of Chicago Press, vol. 117 (3), pp. 417-452.

OECD, (1997). OECD Convention On Combating Bribery Of Foreign Public Officials In International Business Transactions: And Related Documents., Directorate For Financial And Enterprise Affairs.

OTUSANYA, O.J. (2011). The Role Of Multinational Companies In Tax Evasion And Tax Avoidance: The Case Of Nigeria. *Critical Perspectives On Accounting.* Volume 22, Issue 3, pp. 316-332.

OTUSANYA, J.O., (2013). An Investigation Of Corporate Executive Fraud By CEOs In The Nigerian Banking Sector. *African Journal Of Accounting, Auditing And Finance.* Vol. 2, No. 1, pp. 65-89.

PETERSON, R. A., (2000). Constructing Effective Questionnaires. Sage Publications. California.

PETERSON, R. A. & KERIN, R., (1980). Household Income Data Reports In Mail Surveys. *Journal Of Business Research.* 8, pp. 301-313

PHAM, H.T., DAO, T.L. & REILLY, B., (2010). Technical Efficiency In The Vietnamese Manufacturing Sector. *Journal Of International Development.* 22, pp. 503-520.

PHAM, H.T. & REILLY, B., (2009). Ethnic wage inequality in Vietnam. *International Journal of Manpower*, 30 (3). pp. 192-219.

PIERNÉ, G., (2013). Hiring Discrimination Based On National Origin And Religious Closeness: Results From A Field Experiment In The Paris Area. *IZA Journal Of Labor Economics*, Vol. 2, Issue 1, pp. 1-15.

PLATECK, R., PIERRE-PIERRE, F.K. & STEVENS, P., (1985). Development And Design Of Survey Questionnaires. Statistics Canada. Census And Household Survey Methods Division. Ottawa.

PORCANO, T.M., (1988). Correlates Of Tax Evasion. *Journal Of Economic Psychology.* 9, pp. 47-67.

POSNER, D., (2002). The Political Salience Of Cultural Difference: Why Chewas And Tumbukas Are Allies In Zambia And Adversaries In Malawi. Unpublished. UCLA.

Posner, D., (2004). Measuring Ethnic Fractionalization In Africa. *American Journal Of Political Science*. 48, 4.

PriceWaterhouseCoopers, (2010). Nigeria @ 50: Top 50 Tax Issues.

Puhani, P.A., (2000). The Heckman Correction For Selection And Its Critique. *Journal Of Economic Surveys*, Vol. 14, pp.53-68

Rabe-Hesketh, S. & Skrondal, A., (2008). Multilevel And Longitudinal Modeling Using Stata. Stata Corp.

Raiser, M., Ruosso, A. & Steves, F., 2008. Trust In Transition: Cross Country And Firm Evidence. *Journal Of Law, Economics And Organisation*. 24, (2): pp. 407-433.

Ramey, V.A. and Shapiro, M.D., (2001). Displaced Capital: A Study Of Aerospace Plant Closings. *Journal Of Political Economy*. 109, 5, pp. 958-992.

Rand, J. and Tarp, F., (2010). Firm-Level Corruption in Vietnam. United Nations University - World Institute For Development Economics Research, Working Paper No. 2010/16.

Reilly, B. & Witt, R., (2011). Disciplinary Sanctions In English Premiership Football: Is there a racial dimension? *Labour Economics*, 18 (3). pp. 360-370.

Reinikka, R. and Svensson, J., (2003) June. Survey techniques to measure and explain corruption. Policy research working paper series 3071. World Bank

Reinikka, R. & Svensson, J., (2005). Fighting Corruption To Improve Schooling: Evidence From A Newspaper Campaign In Uganda. *Journal Of The European Economic Association*. 3, 2-3, pp. 259-267.

Reinikka, R. and Svensson, J., (2006). Using Micro-Surveys To Measure And Explain Corruption. *World Development*. 34, 2, pp. 359-370.

Rose-Ackerman, S.,(1978). Corruption: A Study In Political Economy. Academic Press. New York.

Rose-Ackerman, S., (1996). Redesigning The State To Fight Corruption: Public Policy For Private Sector. *World Bank Report*.

Rose-Ackerman, S., (1999). Corruption And Government: Causes, Consequences, and Reform. Cambridge University Press. USA.

Rose-Ackerman, (2004). The Challenge Of Poor Governance And Corruption. Mimeo, Copenhagen Consensus And Yale University.

Ross, L., Greene, D. & House, P., (1977). The "False Consensus Effect": An Egocentric Bias In Social Perception And Attribution Processes.*Journal Of Experimental Social Psychology.* 13, pp. 279-301.

Rugg, D., (1941). Experiments In Wording Questions: II. *The Public Opinion Quarterly*, Vol. 5, No. 1, pp. 91-92.

Sah, R., (2007). Corruption Across Countries And Regions: Some Consequences Of Local Osmosis. *Journal Of Economic Dynamics And Control.* Vol. 31(8), pp. 2573-2598.

Shannon, C.E., (1948). A Mathematical Theory Of Communication. *The Bell System Technical Journal.* Vol. 27, pp.379-423, 623-656.

Shaw, P., Katsaiti, M. S. & Jurgilas, M., (2011). Corruption And Growth Under Weak Identification. *Economic Inquiry.* Vol. 49 (1), pp. 264-275

Shleifer, A. & Vishny, R. W., (1993). Corruption. *The Quarterly Journal of Economics.* Vol. 108, No.3, pp. 599-617.

Slonim, R. & Garbarino, E., (2008). Increases In Trust And Atruism From Partner Selection: Experimental Evidence. *Experimental Economics.* 11 (2), pp. 134-153.

Small, K.A. and Hsiao, C., (1985). Multinomial Logit Specification Tests. *International Economic Review*, 26,3, pp. 619-627.

Smith, D.J., (2007). A Culture Of Corruption. Everyday Deception And Popular Discontent In Nigeria. Princeton University Press.

Solomon, J.; Jacobson, S.K.; Wald, K.D. & Gavin, M., (2007). Estimating Illegal Resource Use At A Ugandan Park With The Randomised Response Technique. *Human Dimensions Of Wildlife.* 12, pp. 75-88.

Sood, N., Burger, N., Yoong, J., Kopf, D. & Spreng, C., (2011). Firm-Level Perspectives On Public Sector Engagement With Private Healthcare Providers: Survey Evidence From Ghana And Kenya. Public Library Of Science. 6, 11.

Spicer, M.W. & Lundstedt, S.B., (1976). Understanding Tax Evasion. *Public Finance.* 31, (2), pp. 295-305.

St John, F.A.V.;Keane, A.M.;Edwards-Jones, G.;Jones, L.;Yarnell, R.W.; and Jones, J.P.G., (2011). Identifying Indicators Of Illegal Behaviour: Carnivore Killing In Human-Managed Landscapes. *Proceedings Of The Royal Society B: Biological Sciences*, 279, 1729, pp. 804-812. The Royal Society.

SUDNAM, S. & BRADBURN, N. M., (1974). Response Effects In Surveys: A Review And Synthesis. Aldine Publishing Company. Chicago.

SUDMAN, SEYMOUR; BRADBURN, NORMAN M. and SCHWARZ, NORBERT, (1996). Thinking About Questions: The Application Of Cognitive Processes To Survey Methodology. San Francisco: Jossey-Bass, 1996.

SVENSSON, J., (2002). Who Must Pay Bribes and How Much? Evidence from a Cross Section of Firms. Institute for International Economic Studies.

SVENSSON, J., (2003). Who Must Pay Bribes and How Much? Evidence from a Cross Section of Firms. *The Quarterly Journal of Economics*. 118(1),pp. 207-230

SVENSSON, J.,(2005). Eight Questions About Corruption. *Journal Of Economic Perspectives*, 19(3), 19-42.

SWAMY, A., KNACK, S., LEE, Y. & textscAzfar, O., (2001). Gender And Corruption. *Journal Of Development Economics*. 64,1, pp. 25-55.

TAJFEL, H., BILLIG, M., BUNDY, R.P. & FLAMENT, C., (1971). Social Categorisation And Intergroup Behaviour. *European Journal Of Social Psychology*. 1, pp. 149-178.

TANUR, JUDITH M., (1992). Questions About Questions: Inquiries Into The Cognitive Bases Of Surveys. New York: Russell Sage, 1992.

TAYLOR, C. L. & HUDSON, M. C., 1972. World Handbook Of Political And Social Indicators. 2nd Edition. Yale University Press.

THE CHARTERED INSTITUTE OF TAXATION OF NIGERIA. CITN Nigeria Tax Guide & Statutes.

THE CONVENTION ON BUSINESS INTEGRITY. Tax Laws And Regulations.

The Foreign Corrupt Practices Act. Paragraph 78dd-1, Section (b). (1998) [online at http://www.justice.gov/criminal/fraud/fcpa/docs/fcpa-english.pdf, accessed 17th March 2012]

THOMAS, M., (1999). The Incentive Structure Of Systemic Corruption. Memo. World Bank.

TOURANGEAU, R.; RIPS, L. J. & RASINSKI, K. A., (2000). The Psychology Of Survey Response. Cambridge University Press.

TOURANGEAU, R. & YAN, T., (2007). Sensitive Questions In Surveys. Vol. 133, No.5, pp. 859-883.

Transparency International, (2007). Transparency International's Corruption Perceptions Index. Transparency International, Berlin, Germany.

Treisman, D., (2000). The Causes Of Corruption: A Cross-National Study. *Journal Of Public Economics.* 76 (3), pp. 399-457.

Ufere, N., Perelli, S. Boland, R. & Carlsson, B., (2012). Merchants Of Corruption: How Entrepreneurs Manufacture And Supply Bribes. *World Development.*

United Nations, Department Of Economic And Social Affairs, Statistics Division, (2002). International Standard Industrial Classification Of All Economic Activities (ISIC). Revision 3.1. Statistical Papers: Series M No. 4, Rev 3.1. New York.

United Nations Office On Drugs And Crime, (2005). Economic And Financial Crimes: Challenges To Sustainable Development. The Elecenth United Nations Congress On Crime Prevention And Criminal Justice.

United Nations Office On Drugs And Crime, (2010). Promoting Ethics And Transparency In Business Transactions In Nigeria.

USSR, (1964). Atlas Narodov Mira. The N. N. Miklukho-Maklaya Institute Of Ethnography Of The Academy Of Sciences, Department Of Geodesy And Cartography Of The State Geological Committee Of The USSR.

Rijckeghem, C.V.; Weder, B., (1997). Corruption and the Rate of Temptation-Do Low Wages in the Civil Service Cause Corruption? IMF Working Papers.

Vogel, J., (1974). Taxation And Public Opinion In Sweden: An Interpretation Of Recent Data. National Tax Journal. 27, (4), pp. 499-514.

Warner, S.L., (1965). Randomized Response: A Survey Technique For Eliminating Evasive Answer Bias. *Journal Of The American Statistical Association*, 60 (March): pp. 63-69.

Warwick, D.P. & Lininger, C.A., (1975). The Sample Survey: Theory And Practice. McGraw-Hill. New York.

Wilkinson, S., (2002). Memo On Developing Better Indicators Of Ethnic And Non Ethnic Identities. Presented At LICEP 5th Meeting, Stanford University.

World Bank, (1997). Helping Countries Combat Corruption: The Role Of The World Bank.

WORLD BANK, (2007). Enterprise Survey: Manufacturing Questionnaire.

WORLD BANK, (2009). Enterprise Survey (Productivity And Investment Climate Survey) in Nigeria in 2009. Sample Survey Methodology Design. Etude Economique Conseil

WORLD BANK, (2011). Doing Business 2011. Nigeria. Making A Difference For Entrepreneurs.

WORLD BANK, (2012). Data Access Protocol For Outside Researchers. Available Online: https://www.enterprisesurveys.org/Portal/unprotected/RegisterExternal.aspx?LibId=14. Accessed: 9th September, 2012.

WORLD BANK & INTERNATIONAL FINANCE CORPORATION, (2011). Doing Business 2011: Nigeria. Making A Difference For Entrepreneurs.

ZIMMERMANN, K.F., (1999). Analysis Of Business Surveys, In: *Handbook Of Applied Econometrics. Volume II: Microeconomics*, pp.407-441. Eds: PESARAN, M.H. & SCHMIDT, P..